Landscape Taphonomy Predictably Complicates Demographic Reconstruction

Daniel A. Contreras (corresponding author)
Department of Anthropology, University of Florida
Gainesville, Florida, USA
daniel.contreras@ufl.edu
ORCID: 0000-0002-8127-8789

Brian F. Codding
Department of Anthropology, University of Utah
Salt Lake City, Utah, USA
brian.codding@anthro.utah.edu
ORCID: 0000-0001-7977-8568

Abstract

Accurately reconstructing past human population dynamics is critical for explaining major patterns in the human past. Demand for demographic proxies has driven hopeful interest in the "dates-as-data" approach, which models past demography by assuming a relationship between population size, the production of dateable material, and the corpus of radiocarbon dates produced by archaeological research. However, several biases can affect assemblages of dates, complicating inferences about population size. One serious but potentially addressable issue centers on landscape taphonomy – the ways in which geologic processes structure the preservation and recovery of archaeological sites and/or materials at landscape scales. Here we explore the influence of landscape taphonomy on demographic proxies. More specifically, we evaluate how well demographic proxies may be corrected for taphonomic effects with either a common generalized approach or an empirically-based tailored approach. We demonstrate that frequency distributions of landforms of varying ages can be used to develop local corrections that are more accurate than either global corrections or uncorrected estimates. Using generalized scenarios and a simulated case study based on empirical data on landform ages from the Coso Basin in the western Great Basin region, we illustrate the way in which landscape taphonomy predictably complicates 'dates-as-data' approaches, propose and demonstrate a new method of empirically-based correction, and explore the interpretive ramifications of ignoring or correcting for taphonomic bias.

Keywords: archaeological demography, landscape taphonomy, dates-as-data, radiocarbon SPD

1. Introduction

Accurately reconstructing past human population dynamics is critical for explaining major patterns in the human past, ranging from the development of behavioral modernity (e.g., Powell et al. 2009; Tallavaara et al. 2015; cf. Vaesen et al. 2016) to the emergence and spread of agriculture (e.g., Bevan et al. 2017; Codding et al. 2022; Timpson et al. 2014; Weitzel and Codding 2016). More broadly, demographic proxies are also needed to explain general trends in past human-environment interactions, including human responses to climate change (e.g., Codding et al. 2023; Flohr et al. 2016; Kelly et al. 2013) and the extent and effects past of human land use (e.g., Ellis et al. 2013; Kaplan et al. 2010; Klein Goldewijk et al. 2011). These establish baselines for anthropogenic impacts and inform predictions about future human-climate-land use dynamics (see d'Alpoim Guedes et al. 2016). Past population dynamics are so fundamental that without a reliable method for discerning them, we will be unable to address most of archaeology's "grand challenges" (Kintigh et al. 2014).

Approaches to regional archaeological demography (recently summarized in Drennan et al. 2015) are generally founded upon counts of some class of archaeological feature or artifact whose abundance can be theoretically related to population size. Counts of sites based on archaeological settlement survey are perhaps the simplest and most common proxy. These can be complemented or supplanted by counts of structures or hearths, adjusted by estimates of site area, and fine-tuned to take into account spans of occupation and site function(s). The centrality of archaeological demography, however, has driven hopeful interest in population proxies that are less dependent on systematic archaeological surveys, which are comparatively expensive, slow, and limited in their spatial coverage. Most salient among these over the last two decades has been the "dates-as-data" approach (Rick 1987), which has become the dominant method for reconstructing past population histories (recently, e.g., Bird et al. 2020; Crema and Kobayashi 2020; DiNapoli et al. 2021; Parkinson et al. 2021; Riris 2018; for a recent review see Crema 2022). This method assumes a relationship between population size, the production and survival of dateable material, and the corpus of radiocarbon dates produced by the last ±60 years of archaeological research, and leverages temporal or spatial variation in the distribution of those dates to model past demography.

Methods of demographic reconstruction, like any archaeological endeavor, are fundamentally vulnerable to problems of differential preservation: any population proxy relies on comparing quantities that survive from different time periods, which can for a variety of reasons lead to the underrepresentation of some periods of time and consequent misinterpretations of population dynamics. As a result, estimates of past populations necessarily either assume that all periods are equally represented or attempt to identify which particular periods are underrepresented and apply some estimated correction.

Landscape taphonomy – the ways in which geologic processes structure the preservation and recovery of archaeological sites and/or materials at landscape scales – is one factor that potentially generates systematic bias in demographic reconstruction. This problem is broadly recognized in settlement survey (Banning 2002; Drennan et al. 2015, pp. 162–171; Stafford 1995), and has been recognized since Rick's original dates-as-data paper as one of the factors that attenuates the relationship between a distribution of population over time in a given locale and the assemblage of radiocarbon dates recovered from that region. The most salient attempt at a generalizable solution is Surovell and colleagues' work (Bluhm and Surovell 2019; Surovell et al. 2009; Surovell and Brantingham 2007), which approximates global rates of loss of archaeological material over time by comparing the differences between sedimentary and aerosol (ice core-derived) records of vulcanism; those differences are argued to indicate rates of disappearance of sediments over time. Surovell and colleagues use those approximations to develop a global taphonomic correction, referred to as the "Volcanic" correction (Bluhm and Surovell 2019), which is now widely applied by dates-as-data practitioners (e.g., Barberena et al. 2017; Broughton and Weitzel 2018; Downey et al. 2016; Edinborough et al. 2017; Jones et al. 2021; Peros et al. 2010; Williams 2012) and implemented in the **rcarbon** package as `` `transformSPD` `` (Bevan and Crema 2017).

However, as Surovell and colleagues recognized (2009, p. 1723), deposition and erosion are highly variable in space, and local rates of taphonomic loss can be expected to vary considerably from global ones. This variation will be particularly consequential in regions with active and varied sedimentary histories, leading to systematic biases in demographic reconstructions.

To evaluate the potential bias of local landscape taphonomy, and ways to address it, here we use simulated archaeological data to show that under many taphonomic scenarios neither applying a generalized correction nor ignoring the problem is likely to constitute an adequate response. With a focus on dates-as-date approaches but with results that are broadly applicable to regional archaeological demography, we demonstrate that frequency distributions of landforms of varying ages can be used to develop local corrections that are more accurate than either global corrections or uncorrected estimates.

Using generalized scenarios and a simulated case study based on empirical data on landform ages from the Coso Basin in the western Great Basin region, we illustrate the way in which landscape taphonomy predictably complicates 'dates-as-data' approaches, propose and demonstrate a new method of empirically-based correction, and explore the interpretive ramifications of ignoring or correcting for taphonomic bias.

2. Background

2.1 Landscape Taphonomy

Taphonomic concepts in archaeology most commonly embrace the analysis of post-depositional modification of archaeological materials (Schiffer 1987), but have also been integrated with insights from archaeological survey (e.g., Banning 2002, p. 72) to address regional landscape taphonomy. This can range from regional variation in site formation processes (Borrero 2014) to consideration of the differential survival of sites that are from different time periods and/or located on different landforms (Barton et al. 2002; Burger et al. 2008).

The problem is one that has been most thoroughly discussed in the geoarchaeological literature, in both relatively humid (e.g., Bettis and Benn 1984; Bettis and Mandel 2002; Borejsza et al. 2014; Mandel 2008) and arid (e.g., Fanning et al. 2007; Ravesloot and Waters 2004) environments. These approaches have generally focused on fluvial processes, and particularly the problems posed by destruction or burial of archaeological sites through erosion and deposition. These studies demonstrate that preserved distributions of sites recorded by archaeological surveys of modern land surfaces can be strongly structured by geomorphic patterns as well as by patterns of human settlement and land use. As a result, as Bettis and Mandel conclude, "the accuracy of paleo-demographic…models based on archaeological data depends in large part on the amount and quality of data available for assessing differential temporal and spatial preservation, and regional and local sedimentation rates" (2002: 152). Various cases studies – e.g., the Middle Gila River (Ravesloot and Waters 2004), the Central and Eastern Great Plains (Bettis and Mandel 2002; Mandel 2008), and southern Indiana (Herrmann 2015) – show that both the distribution and the abundance of sites of any given period must be considered in light of the varying ages of extant/exposed landforms in fluvial landscape. The diversity of these examples, as well as modeling of fluvial landscapes (Clevis et al. 2006; Davies et al. 2015), suggests that the problem is pervasive and potentially significant. Ballenger and Mabry (2011) address this with specific reference to the recovery of dateable material used in dates-as-data approaches.

Although fewer case studies address the problem directly in other geomorphic contexts, landscape taphonomy is not limited to fluvial landscapes. For instance, MacInnes and colleagues (2014) address differential availability of landforms for settlement in the Kuril Islands, where landform creation or burial through volcanic processes is the primary process of concern, and Zvelebil and colleagues (1992) consider the impacts on archaeological survey in a southeast Irish landscape of alluviation, sea level change, and peat development. Bailey and Cawthra (2023) review the landscape taphonomic implications of global sea level rise in broad terms. The empirically grounded simulation that we present in Section 3.3 is based on the detailed work on Great Basin landscape taphonomy by Eerkens and colleagues (2007) in the Coso Basin.

For dates-as-data approaches, the role of taphonomy in structuring the distribution of surviving datable material is fundamental. Nevertheless, as Ward and Larcombe (2021) have recently detailed, even if the issue is acknowledged in dates-as-data projects, it is rarely treated in sufficient detail to enable consideration of the likely effects on demographic reconstructions. At best, the vast majority of dates-as-data literature assumes that, all else being equal, older material has been subject to deleterious processes for more time, and is thus less likely to be represented in the archaeological record. Surovell and colleagues (Surovell et al. 2009; Surovell and Brantingham 2007) recognized the importance of this issue, and approximated a solution by developing a "correction" for taphonomic bias using a database of geologic ^{14}C dates associated with volcanic deposits (Bryson et al. 2006) as a measure of the frequency distribution of terrestrial sediments of various ages. They compared this empirical distribution against an independent ice-core-derived aerosol record of Quaternary volcanism, which is unaffected by landscape taphonomy, to produce a global estimate of the impact of taphonomic factors on the survival of terrestrial sediments of different ages. A recent evaluation of the volcanic correction (Bluhm and Surovell 2019) produced largely similar results using an independent set of non-volcanic geologic dates.

While this approach is an ingenious solution to the problem of taphonomic bias, it assumes that local landscape taphonomy mirrors global patterns, smoothing over variation in local surface processes that may produce significant deviations in the post-depositional factors that structure the availability of dateable material in any given region. Since local taphonomy can significantly structure surviving distributions of dateable material, ignoring it can have significant effects on demographic interpretations. Surovell and colleagues (2009, p. 1723) acknowledged this issue and suggested their global correction only as a first approximation. Others (e.g., Attenbrow and Hiscock 2015, p. 32; Rhode et al. 2014, p. 576) also emphasize the importance of attention to local landscape taphonomy and suggest that the appropriateness of a generalized correction should be demonstrated rather than assumed. In spite of this recognition, and although it is clear that in order for summaries of radiocarbon dates to accurately reflect the original distributions of dateable material these taphonomic effects must be accounted for, no systematic approach for dealing with taphonomic effects at local or regional scales exists. Crema's recent (2022) comprehensive review of dates-as-data methods neither explores the magnitude of the problem nor suggests any solutions other than the volcanic correction. Moreover, Surovell and colleagues' global volcanic correction is widely cited (369 citations listed in Google Scholar as of November 2023, though certainly not all of these represent applications of the correction), often without justification of its appropriateness for the region under consideration (though sometimes, e.g., Barberena et al. 2017, with caveats about the applicability of the results).

2.2 "Dates-as-Data" Approaches

Embrace of meta-analysis of archaeological assemblages of ^{14}C dates can be traced to John Rick's (1987) use of ^{14}C dates from the Central Andean preceramic period to argue that ^{14}C dates could be employed as population proxies. Other early efforts can be discerned (see Carleton and Groucutt 2020, p. 2), but Rick's paper is increasingly cited, and its title commonly used to describe this genre of studies.

Following Rick, this "dates-as-data" approach has been founded on the argument that in addition to their traditional role in establishing chronological frameworks for archaeological sites and regions, ^{14}C dates could also figure in analyses of broad demographic patterns in space and time. The central contention is that in spite of various confounding factors, archaeological ^{14}C dates can serve as a population proxy, given an initial assumption that the production of dateable material is roughly proportional to population size at any given time.

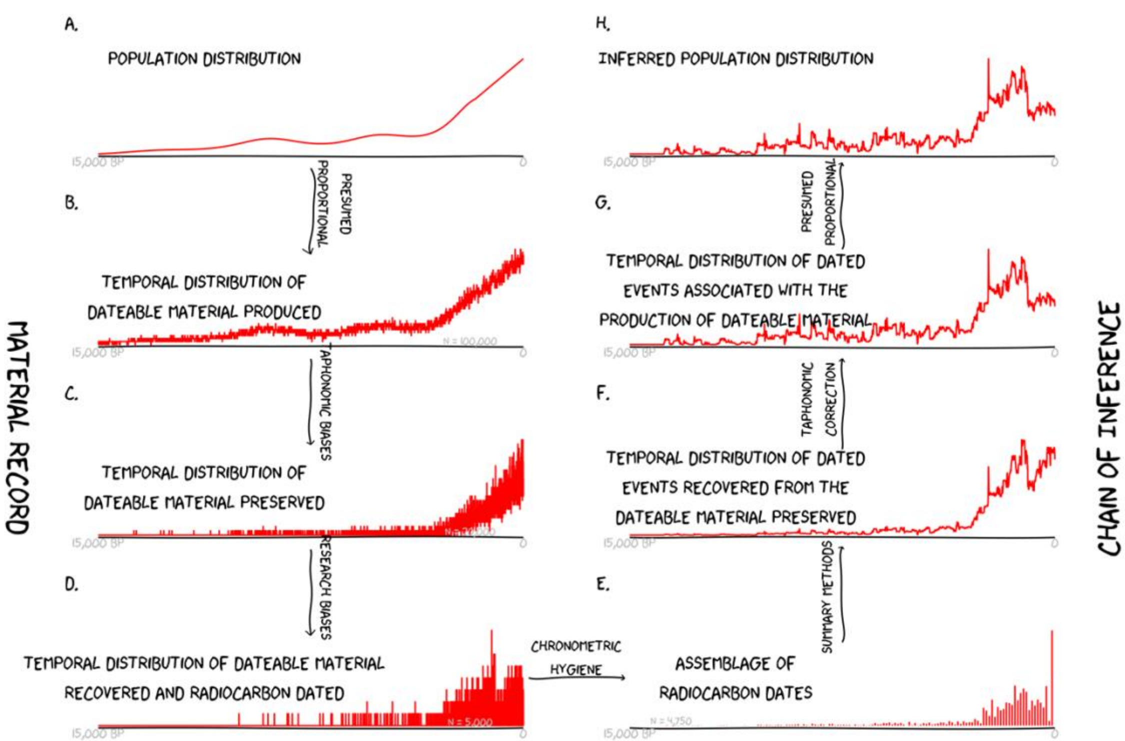

Figure 1: Schematic of the 'dates as data' approach, using simulated data. The creation of the material record (at left) involves the initial production of dateable material and the subsequent transformation of that material by successive processes. A population (A), derived from the Terminal Pleistocene - Holocene estimate produced by Weitzel and Codding (2016), produces dateable material (B) at a rate assumed to be proportional to population. That dateable material is subject to taphonomic processes, which though irregular are cumulative, making older material less likely to be preserved. Here we simulate this taphonomic bias by sampling from the initial distribution with probabilities following the exponential curve described by Surovell and colleagues (2009). The remaining (preserved) dateable material (C) is the population of archaeological material available to be recovered and dated by archaeologists, who for intellectual and budgetary reasons (at

least) do not select material to date at random. The resulting distribution of dateable material (D) is shaped by both the abundance of material available from different periods and the preferential recovery and analysis of material from particular periods. Here we simulate research bias simply by sampling from the preserved distribution with probabilities uniformly equal to 1 for the period before 1000 BP and uniformly equal to .5 for the post-1000 BP period, reflecting the abundance of other dating techniques likely to be used for archaeological material dating to the most recent millennium. The process of making inferences about demographic history from the resulting assemblage of radiocarbon dates (at right) involves the summarization of calibrated radiocarbon dates and then application of a correction for taphonomic effects. After removing 5% of the assemblage to simulate the application of chronometric hygiene to a collection of radiocarbon dates, radiocarbon dates are simulated for each of the remaining calendar dates (using rcarbon::uncalibrate), which are illustrated here with a histogram binning the medians of those radiocarbon dates (E). These simulated radiocarbon dates are calibrated, and then a summed probability distribution (SPD) describing them (F) is calculated. This SPD is taphonomically corrected following the method described by Surovell and colleagues (2009), producing a corrected distribution (G) that is presumed proportional to the population distribution over time (H), and understood as an approximation of (A).

This contention rests upon a) the validity of the relationship between population size and production of dateable material, and b) dismissal of both the effects of research priorities and budgets on the recovery and analysis of dateable material, and the effects of temporally and spatially variable preservation on the ultimate composition of the material record. The latter two processes can significantly structure ^{14}C assemblages in ways that strongly impact interpretation. Addressing these biases, consequently, is vital if "dates as data" approaches are to produce reliable results. We review below the principles and application of the "dates as data" approach, as well as the significant challenges yet to be overcome. These challenges are the product of three assumptions fundamental to the 'dates as data' approach (see Figure 1):

1) past population size is proportional to (∝) the dateable material produced,
2) dateable material produced is proportional to the dateable material now available to sample, and
3) the dateable material now available is representatively sampled.

In order to accurately reconstruct changing populations, archaeologists must develop methods that address whether these assumptions are justifiable for a particular case and, if not, correct for the biases introduced. While this paper focuses on the second fundamental assumption, here we briefly review each as well as the methods developed to try to reduce the impact of biases on dates-as-data. For additional detail we refer the reader to Crema's (2022) recent comprehensive review.

2.2.1 Foundational Assumption 1: Population Size ∝ Dateable Material Produced

The foundational assumption of any attempt to use an assemblage of radiocarbon dates as a population proxy, articulated in Rick's 1987 paper, is clear if not necessarily universally accepted: the production of dateable material at any given time is proportional to population size (Figure 1a and 1b). Rick pointed out from the outset that this relationship was likely to be a function of technology and environment (Rick 1987, p. 57), and argued that the population proxies were only appropriately compared in situations where these were similar, but this caution has not always been observed by subsequent researchers. With the exception of the recent work by Freeman and colleagues (2018), only critiques of "dates as data" approaches (e.g., Attenbrow and Hiscock 2015; Mökkönen 2014; Torfing 2015) tend to raise this issue. Although in principle it is clear that the relationship between population and the production of dateable material may vary over time and/or space, dates-as-data practitioners seem to be content that this risk is either a) unimportant, or b) can be managed by confining analyses to populations within which that relationship is likely to be fairly constant – i.e., where technology and sociopolitical complexity are comparable.

2.2.2 Foundational Assumption 2: Dateable Material Produced ∝ Dateable Material Available

Any approach whose logic relies on diachronic comparison – in the case of 'dates as data' approaches to past population, of the quantities of dateable material produced at different times – must confront the issue of taphonomy (see Section 2.1). Where radiocarbon dates are concerned, the issue is the differential survival of dateable material that might be recovered and analyzed (Figure 1c). However, the "dates as data" literature has generally embraced the convenient assumption that (other things being equal) taphonomic patterns will have a neutral effect on a ^{14}C assemblage, or at least an effect that can be simply corrected.

As early as 1987, however, Rick noted that "preservation processes will discriminate against older dates" (1987, p. 57). Ward and Larcombe (2021, p. 550) have recently reiterated this caution, and a series of studies have explored the potential interpretive ramifications of differential preservation. Ballenger & Mabry (2011) present a case study in which other factors overwhelm production as a determinant of the abundance of dateable material, wherein taphonomic loss cannot be simply modeled ("the conditions that determine preservation/loss have varied through time" [Ballenger and Mabry 2011, p. 1322]). Holdaway and colleagues (2009), on the basis of dates on different kinds of archaeological components in southeastern Australia, and Davies and colleagues (Carney and Davies 2020; Davies et al. 2015), on the basis of model simulations, argue that landscape taphonomy can produce an apparently complex ^{14}C record even if the generative process is simple.

2.2.3 Foundational Assumption 3: Dateable Material Available ∝ Material Dated by Researchers

The issue of research intensity, already recognized in the infancy of "dates as data" approaches by Rick (1987: 57–58), is similarly challenging (Figure 1d). The tacit contention is that an archaeological radiocarbon assemblage can be treated as a random sample of the dateable material produced, in part because disparate research agendas focus on different time periods. A key risk is that research interests and/or budgetary realities may drive research practices: in addition to locally eclectic research preferences, the number of ^{14}C samples dated in any region may best reflect that region's economic fortunes rather than its population in prehistory. Even within regions of comparable prosperity, perceptions as to the relative importance of different archaeological phenomena or periods and the relative utility of ^{14}C and other dating methods mean that resources will be unevenly directed towards dating different periods. Further complicating factors are that researchers collecting ^{14}C results published in academic literature may be unaware of larger and perhaps less selective data sets generated by commercial archaeology (as Crombé and Robinson [2014] observed), and that results may be structured by regional reporting conventions (notable, for instance, in the salience of Wyoming in the Canadian Archaeological Radiocarbon Database [CARD] data [e.g., Chaput et al. 2015, p. Fig. 1; Crema et al. 2017, p. 2]). The effects of even sampling that can be treated as effectively random can also produce patterns that are difficult to distinguish from fluctuations in the abundance of dateable material (Rhode et al. 2014).

2.2.4 "Dates as Data" Methodology

The majority of the "dates as data" literature has focused on the difficulties of summarizing ^{14}C assemblages (see recent reviews in Bronk Ramsey 2017; Crema 2022; Crema and Bevan 2021) and interpreting the resulting summed probability distributions (SPDs); practitioners have generally preferred to take the foundational assumptions for granted (though see Carleton and Groucutt 2021; Freeman et al. 2018).

2.2.4.1 Summarizing Assemblages of ^{14}C dates

Although a few alternatives continue to be explored – e.g., model fits on binned dates (Weitzel and Codding 2016) and summed ranges (Drake et al. 2017) – addressing the uneven probability distributions of calibrated dates by using summed probability distributions has become the dominant method of summarizing ^{14}C assemblages (Figure 1e and 1f), in spite of various methodological and theoretical critiques (e.g., Attenbrow and Hiscock 2015; Bamforth and Grund 2012; Chiverrell et al. 2011; Contreras and Meadows 2014; Culleton 2008; Mökkönen 2014; Torfing 2015). This is likely due in large part to the relative ease with which they can be calculated, coupled with the inability of critiques to suggest a more viable alternative. However, Bronk Ramsey's (2017, pp. 1810–13) discussion of various methods of summarizing ^{14}C dates argues that an adaption of kernel density estimation (KDE) provides a more promising tool for separating signal (date frequency) from noise (effects of the calibration curve and sampling, primarily). Others (e.g., Brown 2015; Codding et al. 2023; Wilson et al. 2023) have explored resampling approaches to explicitly address the uncertainty associated with each radiocarbon date.

2.2.4.2 Correcting for Research Biases

Recent work using Sum distributions to summarize ^{14}C assemblages has in some cases attempted to "correct" ^{14}C assemblages for differential research intensity, by summing the calibrated pooled means of ^{14}C results from individual sites/site-phases (e.g., Buchanan et al. 2008; Shennan and Edinborough 2007; Tallavaara et al. 2010), by summing the calibrated dates for individual sites or site-phases before summing the sums (e.g., Collard et al. 2010; Crema et al. 2016; Hinz et al. 2012; Shennan et al. 2013), or by combining dates from sites (e.g., Balsera et al. 2015), areas (e.g., Goldberg et al. 2016), or site-phases (Timpson et al. 2014) before summing. Chaput and colleagues (2015) use the spatial distribution of the entire assemblage as a measure of the spatial distribution of research, thereby controlling (they argue) for variable intensity of sampling in space, and Crema and colleagues (Crema et al. 2017) address research and other biases by looking for local fluctuations relative to regional trends.

All of these techniques are intended to address the problem of well-funded excavations that produce significantly more ^{14}C dates than other investigations in a region, but they run counter to the fundamental assumption that larger populations would produce more dateable material: pooling gives equal weight to every site or site-phase, thus conflating large and small sites and presuming site populations are static over time. That is, populations of different sizes separated by more than some minimum amount of time are expected to produce different amounts of dateable material, but populations of different sizes separated in space are not. Pooling in this manner leaves unaddressed the question of when the quantity of dates from a particular site, area, or time period represents an anomaly in the amount of research attention paid to that area/site/period, and when it represents a concentration of population. Just as Kent Flannery describes the risk, for a rigid sampling strategy of surface survey in the Basin of Mexico, of missing the metropolis of Teotihuacan (Flannery 1976, p. Ch.5), uniformly binning multiple dates to minimize bias stemming from well-funded investigations may lead "dates as data" researchers to ignore sites that have many dates *specifically because* they are large sites that had large populations.

2.2.4.3 Correcting for Taphonomic Biases

Surovell and colleagues' (2009) work stands out for its creative attempt to confront the issue of taphonomic effects and remains the preferred means of addressing the differential survival of datable material of varying ages (Figure 1g). Although the authors note that their proposed correction is a coarse global approximation and suggest that the best approach would be to develop local corrections for any given study (Bluhm and Surovell 2019, p. 328; Surovell et al. 2009, p. 1723), nevertheless their correction is widely implemented (e.g., Barberena et al. 2017; Broughton and Weitzel 2018; Downey et al. 2016; Edinborough et al. 2017; Fernández-López de Pablo et al. 2019; Zahid et al. 2016), reflecting recognition that taphonomic bias poses a potentially significant problem. However, taphonomic correction is not universally applied (e.g., Codding et al. 2022; Stewart et al. 2021; Tremayne and Winterhalder 2017) and details of correction methods may vary. Williams (2012), for example, preferred a slightly modified version of Surovell and colleagues' empirically-derived equation relating time elapses to survival of material, and argued that either correction produced "unrealistic values for time intervals >25.0 ka" (Williams 2012, p. 584). That dissatisfaction with results that did not match expectations led Williams (2012, p. 586) to argue that "taphonomic correction should not be routinely applied without some discussion of whether time-dependent taphonomic loss is valid as an a priori assumption." Stewart and colleagues (2022, p. 2) make a similar point in more broadly theoretical terms, noting that Surovell's use of a monotonic function to describe taphonomic loss effectively implies that the environmental conditions controlling taphonomic processes were constant over time. Various empirical and simulation studies (e.g., Ballenger and Mabry 2011; Davies et al. 2015; Holdaway et al. 2009; Rhode et al. 2014) – as well as landscape-scale geoarchaeology (see Section 2.1) – demonstrate that in fact taphonomic processes vary in both time and space.

Critiques of Surovell's approach, however, neither argue that taphonomy is unimportant nor suggest any alternative methods of correction. Although Surovell and colleagues explicitly presented their correction as a first approximation in need of further development, and in spite of subsequent cautions about the potentially significant implications of taphonomic effects, only Crema and colleagues' (2017) comparison of local and regional trends has any potential for detecting – much less correcting – taphonomic bias.

2.2.4.4 Interpretation

The interpretation of a corrected distribution of archaeological radiocarbon dates (Figure 1h) represents a final hurdle. Peaks and troughs in summaries of radiocarbon assemblages may result from significant fluctuations in the population that produced the dateable material that survived to be recovered and dated, or they may result from the vagaries of sampling, from the effects of biasing factors, or from unintended effects of methodology (see reviews in Bronk Ramsey 2017; Carleton and Groucutt 2021; Contreras and Meadows 2014; Crema 2022). Slopes – representing rates of change – are similarly vulnerable, particularly over short timespans. The more discerning an interpretation tries to be, the more susceptible it is to confounding factors introduced by taphonomic effects, patterns of research, and simple sampling. Attempts to address challenges of SPD interpretation though methodological improvements – e.g., comparison to growth models (see summary in Crema and Shoda 2021) – tackle the problem of what can be inferred from a summarized ^{14}C assemblage, but do not address how well (or poorly) the sample of ^{14}C dates represents the population for which the SPD is argued to be a proxy.

Both research and taphonomic biases are especially pernicious in that they are spatially and temporally heterogenous, affecting different subsets of large ^{14}C assemblages differently as these biases vary both in space and over time. Interpretations that do not take this variability into account risk overgeneralizing in potentially problematic ways, depending on the questions involved.

3. Simulating Taphonomic Effects and Corrections

As we have detailed above, while taphonomic correction is not entirely standard in dates-as-data approaches, the possibility that older sites are underrepresented has been considered and means of correcting accordingly proposed (Bluhm and Surovell 2019; Surovell et al. 2009; Surovell and Brantingham 2007). Landscape taphonomy has also been identified as a – largely neglected – problem for archaeological assemblages more generally.

The correction developed by Surovell and colleagues (2009) attempts to deal with this by estimating *how much* less likely older material is to survive, and adjusting the SPD accordingly. Their empirically-derived function (Surovell et al. 2009, p. 1717) describes the relationship between time elapsed and probability of survival, positing that for a given age a predictable proportion of material will have survived. As a result, the observed quantity that has survived can be used to estimate how much originally existed by dividing the observed quantity by the expected proportion (Surovell et al. 2009, p. 1718). We mirror this approach here, but addressing the particulars of preservation probabilities for a given assemblage. Specifically, we use simulated data to develop a means of spatially explicit estimation of local taphonomic effects and calculation of corresponding probability weights for ^{14}C samples from different periods. Simulation offers a way to explore the impacts of a) landscapes composed of landforms of varying ages, b) distinct demographic scenarios, and c) various taphonomic corrections.

We consider four scenarios at extremes of these spectra, and explore one empirically grounded realistic scenario based on the Coso Basin in the southwestern Great Basin. For each, we 1) simulate a population and a landscape taphonomic process, 2) produce a simulated sample of radiocarbon dates resulting from the interplay of these factors, and 3) apply dates-as-data methods to attempt to reconstruct the (known) population from which that sample was generated. The results generated in (3) are compared to the simulated population in (1) to explore challenges to demographic reconstruction and the efficacy of different corrections. We implement this approach in the R environment for statistical computing (R Core Team 2021). All code required to replicate our simulations are provided in the Supplementary Material.

3.1 Developing and applying a local taphonomic correction based on landform frequencies

Frequency distributions of landforms of varying ages enable estimation of the varying probabilities of preservation and recovery of archaeological sites of differing ages, and thus estimation of the probabilities of recovering dates from particular age ranges. Using these probabilities to weight dates of different ages in extant ^{14}C assemblage accounts for the differential likelihoods of survival of dateable material produced at varying times, in a process analogous to Surovell's (2009) method but empirically approximating local erosional and depositional processes.

We explore this method by developing a simulation that accounts for:

- production (of sites and dateable material, proportional to population),
- preservation (dependent on both time elapsed and landscape processes – burial and erosion),
- recovery (more possible/likely where landforms that *could* host sites are exposed), and
- reconstruction (of sites/dateable material as a proxy for population).

There are seven steps to the simulation process, summarized below and detailed in the annotated R code included as Supplementary Material.

1. Generate a landform age distribution.
2. Generate a population curve that will provide the probability distribution that governs the sampling in Step 3. This can be derived from a theoretical expectation (e.g., of exponential growth) or from an empirical or hypothetical approximation (e.g., a population reconstruction or inferred trajectory).
3. Use that population curve as a probability distribution governing the selection of a sample of calendar dates over a given span of time at the desired density, adjusting the probabilities according to the frequency distribution of landforms (i.e., sites can only be found on landforms that are at least as old as the sites are) and modeled decay over time (Surovell and Brantingham 2007, p. 1872).
4. Use each of those calendar dates to simulate a radiocarbon date (using, e.g., `R_Simulate` in OxCal [Bronk Ramsey 2009, 2020]; or `uncalibrate` or `unCalibrate` from the **rcarbon** [Bevan and Crema 2017] and **BChron** [Parnell 2015] packages, respectively).
5. Summarize the resulting radiocarbon dates, using, e.g., `rcarbon::spd`. SPDs have been compellingly critiqued as a means of summarizing ^{14}C assemblages (Bronk Ramsey 2017), but remain so common as to be standard.
6. Correct that SPD using both Surovell and colleagues' (2009) volcanic correction and a local correction (derived either directly from the landform distribution in Step 1, or from some empirical approximation). Either correction is applied by dividing the observed value for a given year by the correction-derived proportion expected to have survived for that year.
7. Compare uncorrected, volcanic-corrected, and locally-corrected against the known starting population from Step 2.

In Section 3.2 below, we use this simulation process to explore the reconstruction of known population distributions in both active and stable landscapes. We illustrate the varying success of uncorrected, volcanic-corrected, and locally-corrected SPDs in reconstructing the populations from which these proxies were derived, before considering the implications using a realistic scenario derived from the Coso Basin case study considered by Eerkens and colleagues (2007).

3.2 Simulating Population Scenarios and Geomorphic Extremes

We use notional populations, adjusted for landscape taphonomy and decay, to simulate assemblages of radiocarbon dates that can be subsequently summarized, adjusted for taphonomic effects, and used to approximate the initial population. The correspondence between the reconstructed population and the initial population provides a means of assessing the utility of different approaches to landscape taphonomy (ignoring it, applying a global correction, and applying a local correction) under two scenarios of population growth over time (uniform and logistic) and under two geomorphic scenarios that make it more and less likely that older sites will survive (stable and active environments).

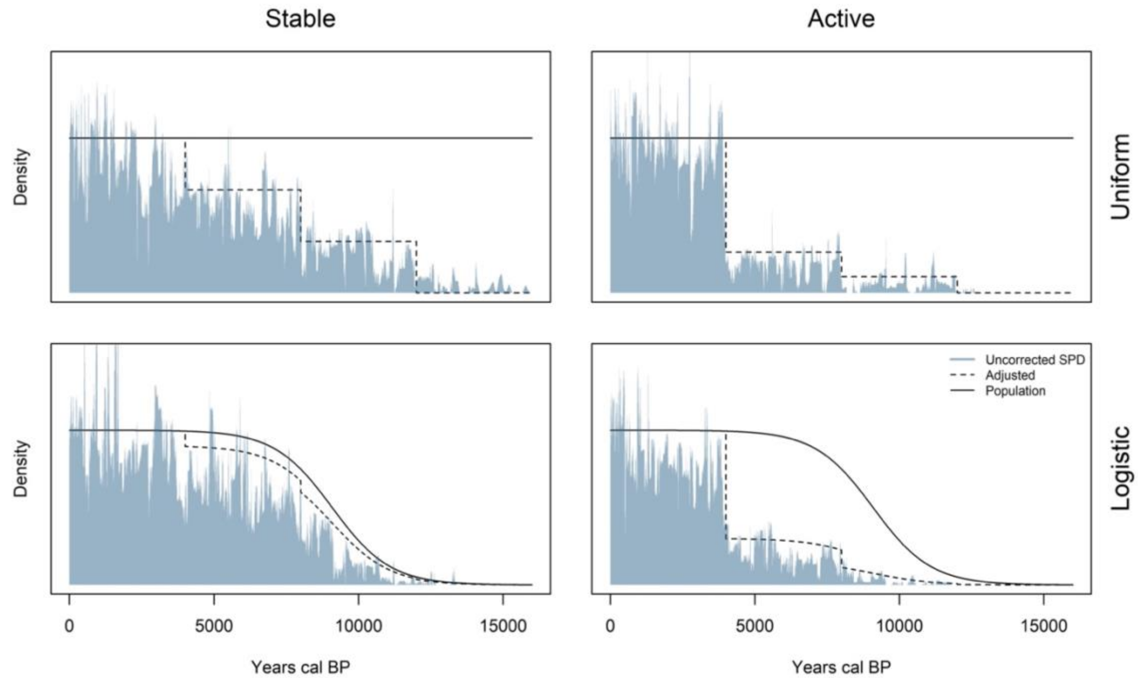

Figure 2: Population/landscape scenarios showing that SPDs reflect geomorphic activity as well as population structure. Solid lines show the original population, dashed lines show the taphonomically-adjusted population, and shaded polygons show the SPD resulting from the sampling (n = 1000) the adjusted population.

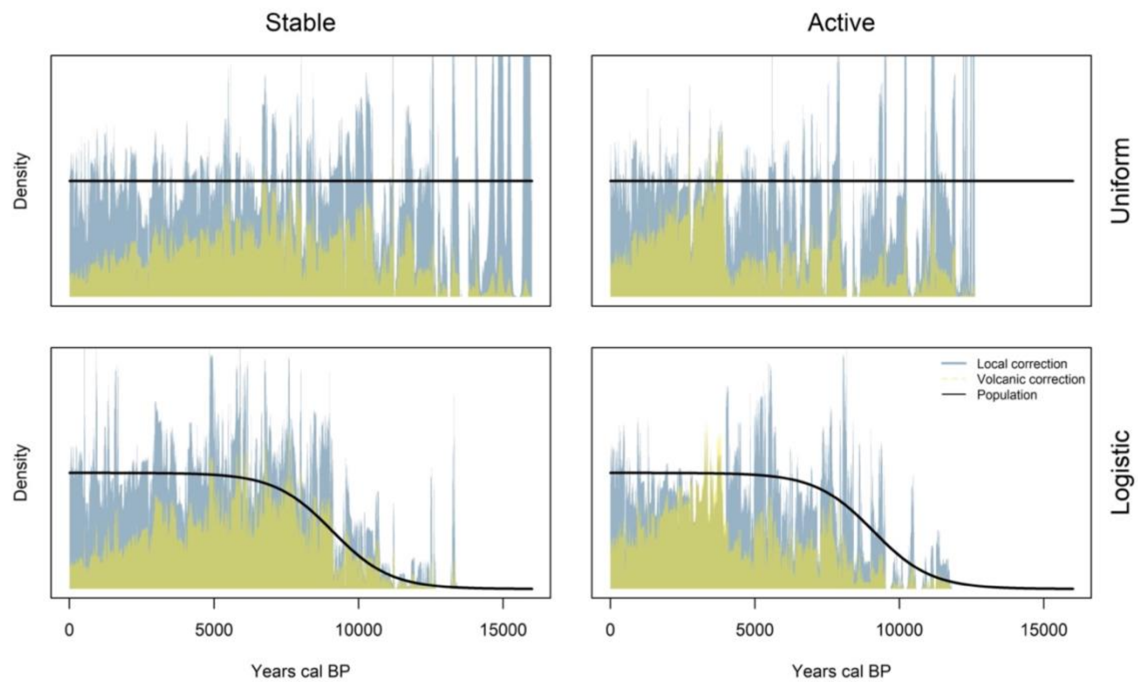

Figure 3: Results of simulations showing the original population (solid line), volcanic-corrected SPD, and locally-corrected SPD under uniform and logistic growth scenarios in stable and active landscapes (all SPDs based on 1000 simulated ^{14}C dates sampled from the landform adjusted population; see Figure 2).

Comparing the uncorrected, volcanic-corrected, and locally-corrected SPDs to the underlying populations from which they are sampled reveals five characteristics of SPDs:

1. Any SPD – corrected or not – is a far from perfect population proxy. The combination of landscape taphonomy, sampling, and calibration introduces significant noise even when an SPD is derived from a uniform distribution. Distinguishing signal from noise remains a fundamental challenge of 'dates as data' approaches. SPDs are best considered like models: all SPDs are wrong; some SPDs are useful (Box 1979).
2. Visual inspection makes clear that for any span of time the locally-corrected SPD (Fig. 3, blue polygon) more closely approximates the underlying population distributions (Fig. 3, black lines) than do the uncorrected SPDs (Fig. 3, yellow polygon) in all four scenarios. While for a few spans of time the volcanic-corrected SPD succeeds as well as the locally-corrected one in approximating the population distribution from which it is derived, for many more spans of time it performs less well.
3. All reconstructions retain artifacts of landscape taphonomy, wherein more active landscapes result in lower probability of recovery of dateable material and hence lower population estimates. All reconstructions do poorly under the conditions of uniform population growth on active landscapes. This is because there is a point where there are so few landforms remaining from which material can be sampled that recovering a sample of sufficient size to accurately estimate the population is very unlikely; the resulting sparseness of samples produces reconstructions that are spiky even when they correct sufficiently that a rolling mean would be high enough to reconstruct the original population.
4. There is greater variance in the locally adjusted SPD than the volcanic-corrected SPD, especially further back in time. This is not surprising as the older dates require greater adjustment, which also amplifies the variance. Future work could further help correct for this by applying a

variance-reducing scaler or smoother and by calculating bootstrapped confidence intervals to focus interpretation on the highest probability region.

5. One limitation of the volcanic-correction is that the calculation implicitly assumes that recent populations are orders of magnitude larger than past ones (see Williams 2012, pp. 584–586). If they are not, more recent estimates will be down-weighted relative to earlier populations, producing population estimates that suggest larger populations in, e.g., the Early Holocene than in 500 BP. We suggest that the best way to handle this is to consider the corrected results *only* for earlier periods, considering instead the *uncorrected* SPD for the recent part of the population distribution. Unfortunately, there is no method, in the abstract, for determining the inflection point – i.e., at what date BP we should stop preferring the corrected results in favor of the original SPD. Some of these issues were recently raised by Bluhm and Surovell (2019).

3.3 A Realistic Coso Basin Simulation

In this section, we use the methods detailed above to simulate a realistic scenario based on Eerkens and colleagues' (2007) study in the Coso Basin. Eerkens and colleagues concluded that the abundance of Early Holocene sites has generally been underestimated due to the extant distribution of landforms of varying ages in the region: the relative scarcity of landforms on which Early Holocene components *could* be present/preserved/found has led to their under-representation in archaeological survey data, and consequently to underestimation of their abundance. That, in turn, has led to reconstructions of site and population densities over time that underestimate the Early Holocene component. In fact, Eerkens and colleagues note, "Early Holocene sites are found throughout the study area *wherever older landforms are present at or near the surface*." (2007, p. 107 [our emphasis]) While Eerkens and colleagues focus on site counts, including as a proxy for population, the issues that they highlight are equally applicable to use of ^{14}C dates as a population proxy. They note: "we believe that site density is a fairly reliable indication of population density. This method of estimating population density avoids many of the problems noted by Surovell and Brantingham (2007), such as tabulating radiocarbon dates." (2007, p. 106)

We draw on the Coso Basin case for 1) frequency distributions of landforms (based on Eerkens et al. 2007: Table 3), and 2) a realistic Holocene population distribution (based on Eerkens et al. 2007: Table 7; we assume for present purposes that Eerkens and colleagues accurately reconstruct Coso Basin populations by accounting for landscape taphonomy). Eerkens and colleagues (2007) exclude the post-1500 BP period from consideration, but Eerkens and Rosenthal (2002, p. 29) consider Coso Basin population growth post-Newberry unlikely; we here follow this in considering post-Newberry population stable. These estimates of relative populations over time provide a realistic population distribution that we use as the basis for this simulated scenario. The point is not the absolute accuracy of the population distribution itself, but rather how well it can be reconstructed from a simulated assemblage of ^{14}C dates that accounts for landscape taphonomy. In this case, that landscape taphonomy is significant: the Coso Basin landscape is one where ~40% of the extant landforms – mid-late Holocene dunes, alluvial fans, and playa deposits – were not available for habitation in the Early Holocene (Table 1).

Table 1: Coso Basin landform frequencies (after Eerkens et al. 2007: Table 3).

Landform	Abbrev	Acreage	Period	Proportion Acreage
pre-Tertiary basement	pTu	1978	Pre- to Early Holocene	0.032
Volcanic rocks	Qv	6530	Pre- to Early Holocene	0.105
Older lakeshore deposits	Qls	34	Pre- to Early Holocene	0.001
Older fan deposits	Qof	22852	Pre- to Early Holocene	0.368

Older lacustrine deposits	Qol	4032	Pre- to Early Holocene	0.065
Older dune sands	Qos	499	Pre- to Early Holocene	0.008
Playa deposits	Qp	2496	Middle to Late Holocene	0.04
Younger fan deposits	Qyf	20989	Middle to Late Holocene	0.338
Dune sands	Qds	2617	Middle to Late Holocene	0.042

We simulate an archaeological radiocarbon assemblage as described in Section 3.1, using landform frequencies and population and population distribution derived from Eerkens and colleagues (2007) as described above. The resulting assemblage of dates is summarized in an SPD, and corrected using both the volcanic correction and a correction derived from the Coso Basin landform frequencies. R code that details this process, like that for the idealized scenarios described in Section 3.2, is available in the Supplementary Material.

Accounting for landscape taphonomy in this context can have significant effects: Figure 4 contrasts an uncorrected SPD of a simulated Coso Basin radiocarbon record (4a) with one adjusted using the global volcanic taphonomic correction suggested by Surovell et al (2009) (4b), and one adjusted using a taphonomic correction factor estimated based on the frequency distribution of landforms in the Coso Basin (4c). Sites that predate 8400 BP cannot be found on mid-late Holocene landforms and have been subject to decay processes for longer, making their survival less likely. As a result, the simulated population distribution (dashed line in Figs. 4a, 4b, and 4c) – moderate in the late Pleistocene and early Holocene, low in the middle Holocene, and relatively high in the later Holocene – produces a distribution of dateable material (solid line in Figs. 4a, 4b, and 4c) that is differentially attenuated over time. Because it is derived from this distribution of surviving material, any radiocarbon-based reconstruction – regardless of the method of summary used – will reflect that pattern, rather than the original population distribution. This is evident in Fig. 4a, where the SPD that summarizes the simulated ^{14}C assemblage (in purple) can be understood as a noisy approximation of the solid line; noise has been introduced by sampling, calibration uncertainty, and summary method.

In fact, the target is the original population distribution, not the distribution of surviving dateable material. In this simulated case study, uncorrected, volcanically/globally corrected, and locally corrected SPDs clearly do not all approximate that original distribution equally well.

The macro-pattern of Coso Basin population – relatively low in the Middle Holocene – is evident in all three results. All three also indicate that the later Early Holocene population was higher than that of the Middle Holocene; the volcanic correction strongly exaggerates this relative high, while the uncorrected SPD slightly underestimates it. The increase in population at the end of the Middle Holocene is apparent in all three results as well, though its magnitude is underestimated by the volcanic correction. As noted above, an unexpected consequence of the volcanic correction is the downward adjustment of the more recent part of a distribution if it is not significantly higher than the earlier portion; in this case the later Holocene is increasingly underestimated by the uncorrected SPD.

An expected consequence of taphonomic decay is that uncorrected and corrected SPDs vary dramatically in the earlier Early Holocene. The uncorrected SPD approximates the relative quantities of surviving dateable material, and as a result appears to indicate a low population that increased slowly over time even though the initial simulated population was stable. Both corrections alleviate this tendency, but they produce very different results, with the local correction much more strongly correcting the Early Holocene. Because the correction only acts upon positive values in the probability distribution (rather than creating data where probabilities are zero), this strong correction increases the variance in the dataset by further exaggerating the positive values. While the result is that the majority of annual estimates vary around the original population distribution, the increased variance due to sampling also produces a noisier signal. Particularly in the earlier Early Holocene and the Terminal Pleistocene, the result gives the impression of boom and bust population cycles. Because the volcanic correction similarly does not attempt to modify probabilities of zero, it also produces a high-variance, noisy pattern of apparent population fluctuation during the same period. Since in this case it is not correcting as strongly, this tendency is less pronounced, but as a result the volcanic correction, like the uncorrected SPD, gives the impression of incremental population growth rather than of substantial and stable population. In addition, even the corrected versions still suggest a contrast between later Early Holocene and earlier Early Holocene, which is in fact the result of the landform distribution (the recent landforms postdate 8400 BP).

In fact, not only is the Early Holocene population underestimated by an SPD; neither correction does enough to recapture the Early Holocene population. This limitation is empirical rather than methodological – correction cannot address an absence of material available to sample (as addressed by Rhode and colleagues [2014]). This problem might be addressed by using a rolling mean or other smoothing approach to capture the central tendency of the corrected SPD, but such an approach requires the tacit assertion that peaks and troughs in the SPD reflect only noise and not signal (i.e., changes in population). Imputing values in the absence of evidence – asserting that for time periods when no archaeological evidence has been found, that absence is due to taphonomic processes and not an absence of occupation – likely is beyond the threshold of correction with which most archaeologists would be comfortable.

This Coso Basin simulation demonstrates that correcting an SPD is likely to be necessary, particularly earlier in the record, that correction may or may not be sufficient to accurately reconstruct a population distribution, and that the choice of correction can significantly affect the results. In this case, archaeologists confronting the distinct population reconstructions would likely infer differing population histories, summarized in Table 2. In the context of Great Basin prehistory, these distinctions have significant interpretive weight. They cast initial colonization and early occupation, responses to mid-Holocene aridity, and Late Holocene re-population in notably different lights, and suggest divergent interpretations of such phenomena as risk management in dynamic environments, adaptive responses to resource uncertainty, and population sensitivity to climate change.

Table 2: Likely inferences about population history of the Coso Basin. Results indicate that local correction would most closely approximate the underlying trends in past populations.

Simulated Population	• stable and moderate in scale throughout the Terminal Pleistocene and Early Holocene • significant Middle Holocene low • rapid growth to a relatively high and stable Late Holocene population
Population Reconstruction	**Likely Inference**
Uncorrected SPD	• consistent population increase throughout the Early Holocene • probable period of population increase immediately preceding a significant mid-Holocene population low
Volcanic correction	• Early Holocene distinguishable into three stages • dramatic late Early Holocene population boom preceding a significant Middle Holocene population low • Middle Holocene low is followed by strong but ephemeral population growth in the Late Holocene
Local correction	• Early Holocene boom-and-bust with generally high population • dramatic Middle Holocene population low • rapid growth to a high and stable Late Holocene population

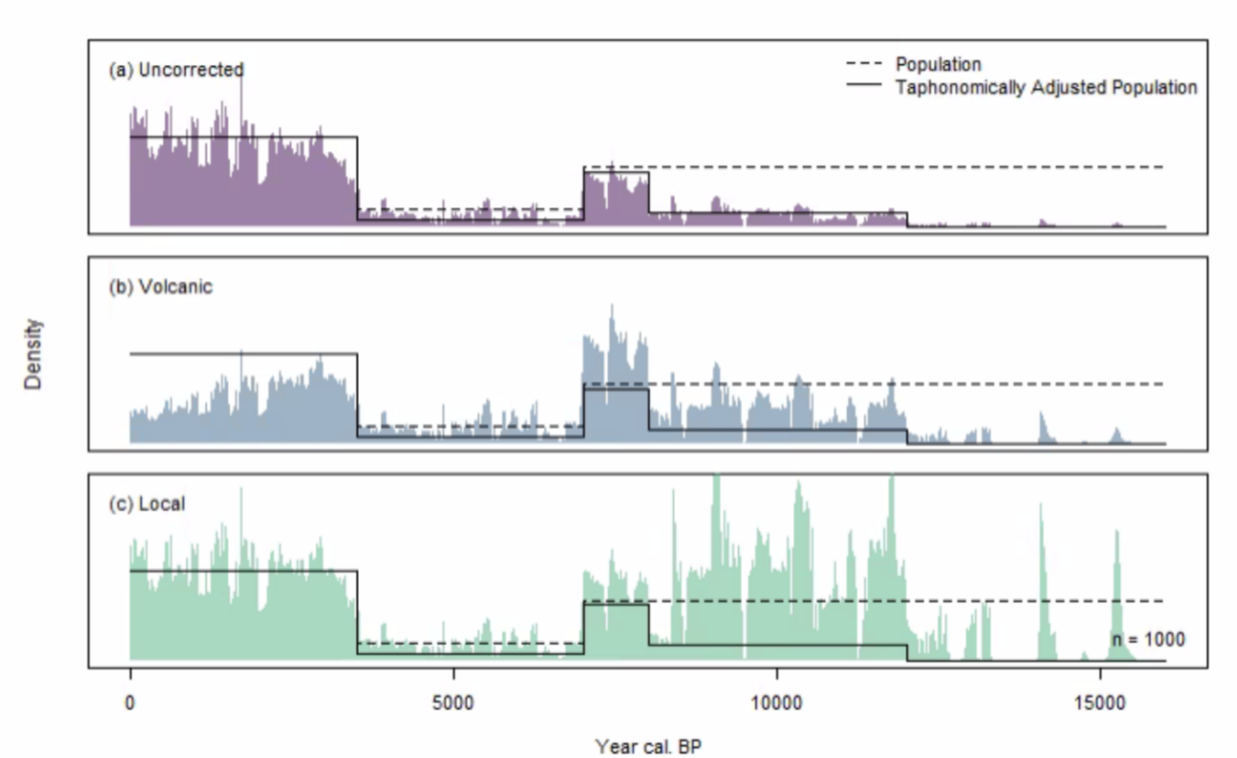

Figure 4: SPDs derived from a simulated Coso Basin radiocarbon assemblage (n=1000): a) uncorrected summed probability distribution, b) summed probability distribution corrected following the global volcanic taphonomic correction produced by Surovell and colleagues (2009), and c) summed probability distribution corrected with a local Coso Basin taphonomic correction. Applying a taphonomic correction at all results in a markedly different distribution (compare a with b or c), and which correction is applied also results in significant changes (compare b and c).The estimated local taphonomic correction employed in (c) is derived from the frequency distribution of landforms of different ages reported by Eerkens and colleagues (2007, p. Table 7), combined with an approximated low rate of taphonomic decay following the exponential curve suggested by Surovell and Brantingham (2007).

4. Discussion

The simulations detailed in Sections 3.2 and 3.3 illustrate some key issues in correcting and interpreting SPDs:

- The volcanic correction doesn't just under-correct or over-correct, but may do one or the other for different spans of time, depending on local landscape taphonomy.
- How much better a local correction approximates the original distribution varies (presumably depending on how much the landform distribution departs from the assumptions of the volcanic correction).
- Further back in time, sparseness of sample leads to underestimates and increased variance with either correction, and neither correction can address an absence of data.
- The volcanic correction down-weights the last 4000 years; this is a mathematical artifact and not an intended effect of the correction.
- One of the problems of any correction that upscales the high values but maintains no-data values (zeroes) is exaggeration of variance, which exacerbates the problem of distinguishing signal from noise.

These issues are fundamental to interpreting assemblages of ^{14}C dates and do not depend on the method used to summarize assemblages of ^{14}C dates. In the terms of the simulations above, methodological improvements focused on improved summary of ^{14}C assemblages (e.g., Bronk Ramsey 2017; Price et al. 2021) may improve how well a taphonomically adjusted population is reconstructed, but remain vulnerable to taphonomic effects. These simulations reveal the significant differences in archaeological interpretation that may result from using different estimates of taphonomic loss (or discounting it), and highlight the importance of selecting the best possible model of taphonomic effects.

The correction suggested by Surovell and colleagues is reasonable and widely employed, but as the Coso Basin simulation demonstrates, it can either over-correct or under-correct, depending on the local landscape history. In the Coso simulation, it over-corrects in the later Early Holocene, but under-corrects in the earlier part of the Early Holocene. These effects result from the relative scarcity of Early Holocene surfaces, which have been buried by later Holocene aeolian, alluvial, and lacustrine deposits. As a result, the volcanic correction, like the uncorrected SPD, underemphasizes the Early Holocene population; for mathematical reasons it also underestimates the Late Holocene population.

Although neither the volcanic correction nor the local Coso Basin correction produces reconstructions that perfectly approximate the initial population distribution, both outperform the uncorrected SPD, and demonstrate the potential significance of taphonomic correction in structuring interpretations of past demography. Although both introduce additional artifacts to data that are already noisy from sampling and calibration effects, the majority of annual estimates from the corrected distributions tend to better approximate initial population distributions. Mismatches between volcanic correction and local landscape history, however, can produce spurious effects, while a local correction – presuming of course that it is accurate – better approximates the initial distribution of dateable material.

Neither method of taphonomic correction is perfect, and neither pretends to address all the potential complications of SPDs. Other systematic biases affecting dated samples – for example, increasing reliance on wood charcoal samples in older contexts where organic preservation can be a significant constraint – can also impact the fidelity with which an SPD reflects population dynamics. A global correction addresses the potentially dramatic under-estimation of older dateable material, but its generalized approach risks over- and under-correcting where local landscape taphonomies diverge from the global average. Local correction avoids this problem, but depends on accurate estimates of the frequency distributions of landforms of different ages. Neither correction can address the uneven sampling that is common if surviving material is sparse (a pattern which might be generated either by relatively small initial quantities and low probability of survival *or* by especially small initial quantities and intensive research [i.e., search for earliest inhabitants]), or if research intensity is heterogenous for different time periods. The magnification of small early signals exacerbates the problem of distinguishing continuous low-level occupation from sporadic occupation (see Rhode et al. 2014), while the successful identification, and accurate and precise dating, of rapid population changes can be critical to interpretation, for example of responses to climate change (Codding et al. 2023).

5. Conclusion

The simulations that we have explored here demonstrate both the potential of summarized radiocarbon data for reconstructing population distributions and the pitfalls of any such approach. It is clear that, as with any archaeological interpretation, biases in the data can significantly structure interpretations, in this case leading to spurious conclusions about past demography. While correcting for taphonomic effects is not a panacea, the structured relationship between frequency distributions of landforms of different ages and distributions of dateable material over time means that biases can be anticipated, described, and accounted for. The results may remain structured by research biases as well as past demography, and the challenge of distinguishing signal from noise will continue to make radiocarbon summaries difficult to interpret (see Figure 4) – but because the complications introduced by taphonomic effects are predictable, they may be accounted for and one source of inaccuracy minimized.

It is clear that taphonomic factors have the potential to skew dates-as-data results. Moreover, and contrary to the assumptions inherent in a global correction, at least some of the likely taphonomic agents

– e.g., sea level change and alluvial deposition and erosion – are likely to have produced taphonomic biases that are heterogeneous both in space and over time. Even within a single study region, a specific taphonomic correction (which no study, so far as we are aware, has attempted to develop) is likely to subsume areas with varying landscape histories, resulting in spatially and temporally distributed over- and under-correction and consequent over- and under-representation of dates and estimation of population. This risk is exacerbated as study regions expand in space and time and more potential diversity is encompassed by an implicit assumption of homogeneity.

As the simulated scenarios discussed here illustrate, without appropriate taphonomic correction, results are likely to be inaccurate, and they are likely to be inaccurate in ways that can meaningfully affect archaeological interpretation. Landscape-scale taphonomic processes are likely to significantly structure the archaeological record, but they are local rather than global. As such, accounting for their effect requires specific attention to local landscape processes. Adjusting summaries of radiocarbon assemblages to account for local/regional frequencies of landforms of varying ages provides an approach that is generalizable to local contexts across the globe. It responds to the as-yet unaddressed appeal that Surovell and colleagues issued when publishing their widely employed global correction: "The ideal approach would be to build local databases of geologic radiocarbon dates that can be used to correct for taphonomic bias, *and to take into account local variation in sedimentation and erosion not captured by the global volcanic model*." (Surovell et al. 2009, p. 1723 [our emphasis])

These simulated cases demonstrate the significant differences in interpretation that may result from using different estimates of taphonomic loss (or discounting it), and highlights the importance of selecting the best possible model of taphonomic effects. Given the potential magnitude of the effects, addressing the differential probabilities of survival of cultural material of different ages is vital to interpretation of regional prehistory and human-environment interactions.

Acknowledgments

This research was supported by National Science Foundation Awards BCS-1921013 and BCS-1921072. The authors thank John Meadows, D. Craig Young, and Duncan Metcalfe for productive conversations about dates-as-data and landscape taphonomy, as well as the various students who have worked on other aspects of this project. We also appreciate the thoughtful and constructive comments provided by two anonymous reviewers.

Declarations

Funding

This research was supported by National Science Foundation Awards 1921013 and 1921072.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Data and Code Availability

The R code used in this project is available in the included Supplementary Information.

References Cited

Attenbrow, V., & Hiscock, P. (2015). Dates and demography: are radiometric dates a robust proxy for long-term prehistoric demographic change? *Archaeology in Oceania*, *50*(2), 29–35.

Bailey, G., & Cawthra, H. C. (2023). The significance of sea-level change and ancient submerged landscapes in human dispersal and development: A geoarchaeological perspective. *Oceanologia* *65*(1), 50-70.

Ballenger, J. A. M., & Mabry, J. B. (2011). Temporal frequency distributions of alluvium in the American Southwest: taphonomic, paleohydraulic, and demographic implications. *Journal of Archaeological Science*, *38*(6), 1314–1325.

Balsera, V., Díaz-del-Río, P., Gilman, A., Uriarte, A., & Vicent, J. M. (2015). Approaching the demography of late prehistoric Iberia through summed calibrated date probability distributions (7000 - 2000 cal BC). *Quaternary International*, *386*(C), 208–211.

Bamforth, D. B., & Grund, B. (2012). Radiocarbon calibration curves, summed probability distributions, and early Paleoindian population trends in North America. *Journal of Archaeological Science*, *39*(6), 1768–1774.

Banning, E. B. (2002). *Archaeological Survey*. New York: Springer Science + Business Media.

Barberena, R., Méndez, C., & de Porras, M. E. (2017). Zooming out from archaeological discontinuities: The meaning of mid-Holocene temporal troughs in South American deserts. *Journal of Anthropological Archaeology*, *46*, 68–81.

Barton, C. M., Bernabeu, J., Aura, J. E., Garcia, O., & La Roca, N. (2002). Dynamic landscapes, artifact taphonomy, and landuse modeling in the western Mediterranean. *Geoarchaeology: An International Journal*, *17*(2), 155–190.

Bettis, E. A., & Benn, D. W. (1984). An archaeological and geomorphological survey in the central des moines river valley, iowa. *Plains Anthropologist*, *29*(105), 211–227.

Bettis, E. A., & Mandel, R. D. (2002). The effects of temporal and spatial patterns of Holocene erosion and alluviation on the archaeological record of the Central and Eastern Great Plains, U.S.A. *Geoarchaeology*, *17*(2), 141–154. https://doi.org/10.1002/gea.10006

Bevan, A., Colledge, S., Fuller, D., Fyfe, R., Shennan, S., & Stevens, C. (2017). Holocene fluctuations in human population demonstrate repeated links to food production and climate. *Proceedings of the National Academy of Sciences*, 114(49), E10524-E10531.

Bevan, A., & Crema, E. R. (2017). *rcarbon: Methods for calibrating and analysing radiocarbon dates*. https://CRAN.R-project.org/package=rcarbon

Bird, D., Freeman, J., Robinson, E., Maughan, G., Finley, J. B., Lambert, P. M., & Kelly, R. L. (2020). A first empirical analysis of population stability in North America using radiocarbon records. *The Holocene*, *30*(9), 1345–1359.

Bluhm, L. E., & Surovell, T. A. (2019). Validation of a global model of taphonomic bias using geologic radiocarbon ages. *Quaternary Research*, 91(1), 325–328.

Borejsza, A., Frederick, C., Alatorre, L. M., & Joyce, A. (2014). Alluvial stratigraphy and the search for preceramic open-air sites in highland Mesoamerica. *Latin American Antiquity*, *25*(3), 278–299.

Borrero, L. A. (2014). Regional Taphonomy. In C. Smith (Ed.), *Encyclopedia of Global Archaeology*. Springer.

Box, G. E. (1979). Robustness in the strategy of scientific model building. In R. L. Launer & G. N. Wilkinson (Eds.), *Robustness in Statistics* (Vol. 1, pp. 201–236). New York: Academic Press.

Bronk Ramsey, C. (2009). Bayesian analysis of radiocarbon dates. *Radiocarbon*, *51*(1), 337–360.

Bronk Ramsey, C. (2017). Methods for Summarizing Radiocarbon Datasets. *Radiocarbon*, *59*(6), 1809–1833.

Bronk Ramsey, C. (2020). OxCal. Oxford. https://c14.arch.ox.ac.uk/oxcal.html

Broughton, J. M., & Weitzel, E. M. (2018). Population reconstructions for humans and megafauna suggest mixed causes for North American Pleistocene extinctions. *Nature Communications*, 1–13.

Brown, W. A. (2015). Through a filter, darkly: Population size estimation, systematic error, and random error in radiocarbon-supported demographic temporal frequency analysis. *Journal of Archaeological Science*, *53*, 133–147.

Bryson, R., Bryson, R., & Ruter, A. (2006). A calibrated radiocarbon database of late Quaternary volcanic eruptions. *Earth Discussions*, *1*(2), 123–134.

Buchanan, B., Collard, M., & Edinborough, K. (2008). Paleoindian demography and the extraterrestrial impact hypothesis. *Proceedings of the National Academy of Sciences*, *105*(33), 11651–11654.

Burger, O., Todd, L. C., & Burnett, P. (2008). The behavior of surface artifacts: Building a landscape taphonomy on the High Plains. In L. L. Scheiber & B. J. Clark (Eds.), *Archaeological Landscapes on the High Plains* (pp. 203–236). Boulder: University Press of Colorado.
Carleton, W. C., & Groucutt, H. S. (2021). Sum things are not what they seem: Problems with point-wise interpretations and quantitative analyses of proxies based on aggregated radiocarbon dates. *The Holocene* 31(4), 630-643.
Carney, M., & Davies, B. (2020). Agent-based modeling, scientific reproducibility, and taphonomy: A successful model implementation case study. *Journal of Computer Applications in Archaeology*, *3*(1), 182–196.
Chaput, M. A., Kriesche, B., Betts, M., Martindale, A., Kulik, R., Schmidt, V., & Gajewski, K. (2015). Spatiotemporal distribution of Holocene populations in North America. *Proceedings of the National Academy of Sciences*, *112*(39), 12127–12132.
Chiverrell, R. C., Thorndycraft, V. R., & Hoffmann, T. O. (2011). Cumulative probability functions and their role in evaluating the chronology of geomorphological events during the Holocene. *Journal of Quaternary Science*, *26*(1), 76–85.
Clevis, Q., Tucker, G. E., Lock, G., Lancaster, S. T., Gasparini, N., Desitter, A., & Bras, R. L. (2006). Geoarchaeological simulation of meandering river deposits and settlement distributions: A three-dimensional approach. *Geoarchaeology*, *21*(8), 843–874.
Codding, B. F., Brenner Coltrain, J., Louderback, L., Vernon, K. B., Magargal, K. E., Yaworsky, P. M., et al. (2022). Socioecological Dynamics Structuring the Spread of Farming in the North American Basin-Plateau Region. *Environmental Archaeology*, *27*(4), 434–446. https://doi.org/10.1080/14614103.2021.1927480
Codding, B. F., Roberts, H., Eckerle, W., Brewer, S. C., Medina, I. D., Vernon, K. B., & Spangler, J. S. (2023). Can we reliably detect adaptive responses of hunter-gatherers to past climate change? Examining the impact of Mid-Holocene drought on Archaic settlement in the Basin-Plateau Region of North America. *Quaternary International*.
Collard, M., Edinborough, K., Shennan, S., & Thomas, M. G. (2010). Radiocarbon evidence indicates that migrants introduced farming to Britain. *Journal of Archaeological Science*, *37*(4), 866–870.
Contreras, D. A., & Meadows, J. (2014). Summed radiocarbon calibrations as a population proxy: a critical evaluation using a realistic simulation approach. *Journal of Archaeological Science*, *52*, 591–608.
Crema, E. R. (2022). Statistical inference of prehistoric demography from frequency distributions of radiocarbon dates: A review and a guide for the perplexed. *Journal of Archaeological Method and Theory*, 29(4):1387-1418.
Crema, E. R., & Bevan, A. (2021). Inference from large sets of radiocarbon dates: software and methods. *Radiocarbon*, *63*(1), 23–39.
Crema, E. R., Bevan, A., & Shennan, S. (2017). Spatio-Temporal Approaches to Archaeological Radiocarbon Dates. *Journal of Archaeological Science*, *87*, 1–9.
Crema, E. R., Habu, J., Kobayashi, K., & Madella, M. (2016). Summed probability distribution of ^{14}C dates suggests regional divergences in the population dynamics of the Jomon Period in eastern Japan. *PLoS ONE*, *11*(4), e0154809.
Crema, E. R., & 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., & 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.
Crombé, P., & Robinson, E. (2014). ^{14}C Dates as demographic proxies in Neolithisation models of northwestern Europe: a critical assessment using Belgium and northeast France as a case-study. *Journal of Archaeological Science*, *52*, 558–566.

Culleton, B. J. (2008). Crude demographic proxy reveals nothing about Paleoindian population. *Proceedings of the National Academy of Sciences*, *105*(50), E111.

d'Alpoim Guedes, J. A., Crabtree, S. A., Bocinsky, R. K., & Kohler, T. A. (2016). Twenty-first century approaches to ancient problems: Climate and society. *Proceedings of the National Academy of Sciences*, *113*(51), 14483–14491.

Davies, B., Holdaway, S. J., & Fanning, P. C. (2015). Modelling the palimpsest: An exploratory agent-based model of surface archaeological deposit formation in a fluvial arid Australian landscape. *The Holocene*, *26*(3), 450–463.

DiNapoli, R., Crema, E., Lipo, C., Rieth, T., & Hunt, T. (2021). Approximate Bayesian Computation of radiocarbon and paleoenvironmental record shows population resilience on Rapa Nui (Easter Island). *Nature Communications*, *12*(1), 3939.

Downey, S. S., Haas, Jr., W. R., & Shennan, S. J. (2016). European Neolithic societies showed early warning signals of population collapse. *Proceedings of the National Academy of Sciences*, *113*(35), 9751–9756.

Drake, B. L., Blanco-González, A., & Lillios, K. T. (2017). Regional demographic dynamics in the Neolithic transition in Iberia: Results from summed calibrated date analysis. *Journal of Archaeological Method and Theory*, *24*(3), 796–812.

Drennan, R. D., Berrey, C. A., & Peterson, C. E. (2015). *Regional Settlement Demography in Archaeology*. Clinton Corners, NY: Eliot Werner Publications.

Edinborough, K., Porčić, M., Martindale, A., Brown, T. J., Supernant, K., & Ames, K. M. (2017). Radiocarbon test for demographic events in written and oral history. *Proceedings of the National Academy of Sciences*, *114*(47), 12436–12441.

Eerkens, J. W., & Rosenthal, J. S. (2002). Transition from geophyte to seed processing: Evidence for intensification from thermal features near China Lake, northern Mojave Desert. *Pacific Coast Archaeological Society Quarterly*, *38*(2–3), 19–36.

Eerkens, J. W., Rosenthal, J. S., Young, D. C., & King, J. (2007). Early Holocene Landscape Archaeology in the Coso Basin, Northwestern Mojave Desert, California. *North American Archaeologist*, *28*(2), 87–112.

Ellis, E. C., Kaplan, J. O., Fuller, D. Q., Vavrus, S., Goldewijk, K. K., & Verburg, P. H. (2013). Used planet: A global history. *Proceedings of the National Academy of Sciences*, *110*(20), 7978–7985.

Fanning, P. C., Holdaway, S. J., & Rhodes, E. J. (2007). A geomorphic framework for understanding the surface archaeological record in arid environments. *Geodinamica Acta*, *20*(4), 275–286. https://doi.org/10.3166/ga.20.275-286

Fernández-López de Pablo, J., Gutiérrez-Roig, M., Gómez-Puche, M., McLaughlin, R., Silva, F., & Lozano, S. (2019). Palaeodemographic modelling supports a population bottleneck during the Pleistocene-Holocene transition in Iberia. *Nature Communications*, *10*(1), 1872.

Flannery, K. V. (Ed.). (1976). *The Early Mesoamerican Village*. New York: Academic Press.

Flohr, P., Fleitmann, D., Matthews, R., Matthews, W., & Black, S. (2016). Evidence of resilience to past climate change in Southwest Asia: Early farming communities and the 9.2 and 8.2 ka events. *Quaternary Science Reviews*, *136*(C), 23–39.

Freeman, J., Byers, D. A., Robinson, E., & Kelly, R. L. (2018). Culture Process and the Interpretation of Radiocarbon Data. *Radiocarbon*, 60(2), 453–467.

Goldberg, A., Mychajliw, A. M., & Hadly, E. A. (2016). Post-invasion demography of prehistoric humans in South America. *Nature*, *532*(7598), 232–235.

Herrmann, E. W. (2015). How bedrock-controlled channel migration can structure selective preservation of archaeological sites: Implications for modeling Paleoindian settlement. *Geoarchaeology*, *31*(1), 58–74.

Hinz, M., Feeser, I., Sjögren, K.-G., & Müller, J. (2012). Demography and the intensity of cultural activities: an evaluation of Funnel Beaker Societies (4200-2800 cal BC). *Journal of Archaeological Science*, *39*(10), 3331–3340.

Holdaway, S. J., Fanning, P. C., & Littleton, J. (2009). Assessing the frequency distribution of radiocarbon determinations from the archaeological record of the Late Holocene in western NSW, Australia. In A. S. Fairbairn, S. O'Connor, & B. Marwick (Eds.), *New Directions in Archaeological Science* (pp. 1–11). Canberra: ANU E Press.

Jones, T. L., Coltrain, J. B., Jacobs, D. K., Porcasi, J., Brewer, S. C., Buckner, J. C., et al. (2021). Causes and consequences of the late Holocene extinction of the marine flightless duck (Chendytes lawi) in the northeastern Pacific. *Quaternary Science Reviews*, *260*, 106914.

Kaplan, J. O., Krumhardt, K. M., Ellis, E. C., Ruddiman, W. F., Lemmen, C., & Klein Goldewijk, K. (2010). Holocene carbon emissions as a result of anthropogenic land cover change. *The Holocene*, *21*(5), 775–791. https://doi.org/10.1177/0959683610386983

Kelly, R. L., Surovell, T. A., Shuman, B. N., & Smith, G. M. (2013). A continuous climatic impact on Holocene human population in the Rocky Mountains. *Proceedings of the National Academy of Sciences*, *110*(2), 443–447.

Kintigh, K. W., Altschul, J. H., Beaudry, M. C., Drennan, R. D., Kinzig, A. P., Kohler, T. A., et al. (2014). Grand challenges for archaeology. *Proceedings of the National Academy of Sciences*, *111*(3), 879–880.

Klein Goldewijk, K., Beusen, A., Van Drecht, G., & De Vos, M. (2011). The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecology and Biogeography*, *20*(1), 73–86.

MacInnes, B., Fitzhugh, B., & Holman, D. (2014). Controlling for landform age when determining the settlement history of the Kuril Islands. *Geoarchaeology*, *29*(3), 185–201.

Mandel, R. D. (2008). Buried Paleoindian-age landscapes in stream valleys of the central plains, USA. *Geomorphology*, *101*(1), 342–361. https://doi.org/10.1016/j.geomorph.2008.05.031

Mökkönen, T. (2014). Archaeological radiocarbon dates as a population proxy: skeptical view. *Fennoscandia Archaeologica*, *31*, 125–134.

Parkinson, E. W., McLaughlin, T. R., Esposito, C., Stoddart, S., & Malone, C. (2021). Radiocarbon dated trends and central Mediterranean prehistory. *Journal of World Prehistory*, *34*(3), 317–379.

Parnell, A. (2015). *Bchron: Radiocarbon dating, age-depth modelling, relative sea level rate estimation, and non-parametric phase modelling*. https://CRAN.R-project.org/package=Bchron

Peros, M. C., Munoz, S. E., Gajewski, K., & Viau, A. E. (2010). Prehistoric demography of North America inferred from radiocarbon data. *Journal of Archaeological Science*, *37*(3), 656–664. https://doi.org/10.1016/j.jas.2009.10.029

Powell, A., Shennan, S., & Thomas, M. G. (2009). Late Pleistocene demography and the appearance of modern human behavior. *Science*, *324*(5932), 1298–1301.

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

R Core Team. (2021). *R: A language and environment for statistical computing* (manual). Vienna, Austria. https://www.R-project.org/

Ravesloot, J. C., & Waters, M. R. (2004). Geoarchaeology and archaeological site patterning on the middle Gila River, Arizona. *Journal of Field Archaeology*, *29*(1–2), 203–214. https://doi.org/10.1179/jfa.2004.29.1-2.203

Rhode, D., Brantingham, P. J., Perreault, C., & Madsen, D. B. (2014). Mind the gaps: testing for hiatuses in regional radiocarbon date sequences. *Journal of Archaeological Science*, *52*, 567–577.

Rick, J. W. (1987). Dates as data: an examination of the Peruvian preceramic radiocarbon record. *American Antiquity*, *52*(1), 55–73.

Riris, P. (2018). Dates as data revisited: A statistical examination of the Peruvian preceramic radiocarbon record. *Journal of Archaeological Science*, *97*, 67–76.

Schiffer, M. B. (1987). *Formation Processes of the Archaeological Record*. Albuquerque, N.M.: University of New Mexico Press.

Shennan, S., & Edinborough, K. (2007). Prehistoric population history: from the Late Glacial to the Late Neolithic in central and northern Europe. *Journal of Archaeological Science*, *34*(8), 1339–1345.

Shennan, S., Timpson, A., Edinborough, K., Colledge, S. M., Kerig, T., Manning, K., et al. (2013). Regional population collapse followed initial agriculture booms in mid-Holocene Europe. *Nature Communications*, *4*, 1–8.

Stafford, C. (1995). Geoarchaeological perspectives on paleolandscapes and regional subsurface archaeology. *Journal of Archaeological Method and Theory*, *2*(1), 69–104.

Stewart, M., Carleton, W. C., & Groucutt, H. S. (2021). Climate change, not human population growth, correlates with Late Quaternary megafauna declines in North America. *Nature Communications*, *12*(1), 965.

Stewart, M., Carleton, W. C., & Groucutt, H. S. (2022). Reply to: Accurate population proxies do not exist between 11.7 and 15 ka in North America. *Nature Communications*, *13*(1), 4693. https://doi.org/10.1038/s41467-022-32356-3

Surovell, T. A., & Brantingham, P. J. (2007). A note on the use of temporal frequency distributions in studies of prehistoric demography. *Journal of Archaeological Science*, *34*(11), 1868–1877.

Surovell, T. A., Finley, J. B., Smith, G. M., Brantingham, P. J., & Kelly, R. L. (2009). Correcting temporal frequency distributions for taphonomic bias. *Journal of Archaeological Science*, *36*(8), 1715–1724.

Tallavaara, M., Luoto, M., Korhonen, N., Järvinen, H., & Seppä, H. (2015). Human population dynamics in Europe over the last glacial maximum. *Proceedings of the National Academy of Sciences*, *112*(27), 8232–8237.

Tallavaara, M., Pesonen, P., & Oinonen, M. (2010). Prehistoric population history in eastern Fennoscandia. *Journal of Archaeological Science*, *37*(2), 251–260.

Timpson, A., Colledge, S., Crema, E., Edinborough, K., Kerig, T., Manning, K., et al. (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*(C), 549–557.

Torfing, T. (2015). Neolithic population and summed probability distribution of ^{14}C-dates. *Journal of Archaeological Science*, *63*, 193–198.

Tremayne, A. H., & Winterhalder, B. (2017). Large mammal biomass predicts the changing distribution of hunter-gatherer settlements in mid-late Holocene Alaska. *Journal of Anthropological Archaeology*, *45*, 81–97.

Vaesen, K., Collard, M., Cosgrove, R., & Roebroeks, W. (2016). Population size does not explain past changes in cultural complexity. *Proceedings of the National Academy of Sciences*, *113*(16), E2241–E2247.

Ward, I., & Larcombe, P. (2021). Sedimentary unknowns constrain the current use of frequency analysis of radiocarbon data sets in forming regional models of demographic change. *Geoarchaeology*, *36*(3), 546–570.

Weitzel, E. M., & Codding, B. F. (2016). Population growth as a driver of initial domestication in Eastern North America. *Royal Society Open Science*, *3*(8), 160319.

Williams, A. N. (2012). The use of summed radiocarbon probability distributions in archaeology: a review of methods. *Journal of Archaeological Science*, *39*(3), 578–589.

Wilson, K. M., McCool, W. C., & Coltrain, J. B. (2023). Climate and oceanic condition changes influence subsistence economic adaptation through intensification on the Central Andean coasts. *Quaternary International*.

Zahid, H. J., Robinson, E., & Kelly, R. L. (2016). Agriculture, population growth, and statistical analysis of the radiocarbon record. *Proceedings of the National Academy of Sciences*, *113*(4), 931–935.

Zvelebil, M., Green, S. W., & Macklin, M. G. (1992). Archaeological landscapes, lithic scatters, and human behavior. In J. Rossignol & L. Wandsnider (Eds.), *Space, time, and archaeological landscapes* (pp. 193–226). New York: Springer.