Inferring Population Dynamics from Multiple Archaeological Proxies: an Overview of Methods and Challenges

(Article begins on next page)

Inferring Population Dynamics from Multiple Archaeological Proxies: an Overview of Methods and Challenges

Alessio Palmisano (Department of Historical Studies, University of Turin)

1. Introduction

Nowadays, population growth occupies a central role in the public debate for its implications on subsistence strategy, environmental change, and its relationship with exogenous factors such as climate change. A recent report by the UN predicts an average annual rate of global population growth of less than 1% from 2021, which should gradually approach 0.1 % by the end of the 21st century[1]. Of course, modern nations and supra-national organizations need to monitor the population and its variation in order to take decisions in terms of infrastructure planning, food and water resources management, fiscal policies, etc. This is possible via a detailed census of the population that in most countries is carried out every 10 years. Nevertheless, in most countries of the world detailed census as we know it today started taking place only in the second half of the 20th century with the exception of some countries such as the UK, where the modern census was introduced in 1801 following the influence of Malthus after his *An Essay on the Principle of Population*[2]. Therefore, in the absence of reliable census data from the past, archaeology is the only discipline among the social sciences, which can provide a picture of population dynamics over the *longue durée* in pre-modern societies. This allows archaeologists to build narratives on population fluctuations and their implications for time spans covering thousands of years and not only the past two centuries. In the past two decades there has been a renewed interest in palaeodemography and archaeology has been leading this research agenda in the light of novel innovative approaches and a large amount of digital data freely available online[3].

Past studies have emphasized the role of population size as a driver for cultural change[4], agricultural innovations and subsistence strategies[5], social complexity[6], socioeconomic outputs[7], shifts in settlement systems[8], and intra-group competition[9]. More recently, some studies have emphasized how more favourable climatic conditions with the onset of the Holocene coupled

---

[1] United Nations, Department of Economic and Social Affairs, Population Division (2022). World Population Prospects 2022, Online Edition.
[2] T. R. Malthus, An Essay on the Principle of Population (London 1798).
[3] J. C. French – P. Riris – J. Fernandez-Lopez de Pablo – S. Lozano – F. Silva, A manifesto for palaeodemography in the twenty-first century. Philosophical Transactions of the Royal Society B 376, 2021.
[4] R. Naroll, A preliminary index of social development, Am. Anthropol. 58 (4), 1956.
[5] E. Boserup, The Condition of Agricultural Growth. The Economics of Agrarian Change under Population Pressure (London 1965); S. Shennan, Population, culture history, and the dynamics of culture change, Curr. Anthropol. 41, 2000.
[6] A.W. Johnson - T.K. Earle, The Evolution of Human Societies: from Foraging Group to Agrarian State (Stanford 2000); G. M. Feinman, Size, complexity, and organizational variation: a comparative approach, Cross Cult. Res. 45 (1), 2011, 37-58; P. Turchin – T. Currie, T. – H. Whitehouse et al., Quantitative historical analyses uncover a single dimension of complexity that structures global variation in human social organization. *Proceedings of the National Academy of Sciences 115*(2), 2018, 144–151.
[7] L.M.A. Bettencourt - J. Lobo - D. Helbing - C. Kuhnert - G.B. West, Growth, innovation, scaling, and the pace of life in cities, Proc. Natl. Acad. Sci. Unit. States Am. 104 (17), 2007, 7301-7306; M. Altaweel - A. Palmisano, Urban and transport scaling: northern Mesopotamia in the late chalcolithic and Bronze age, J. Archaeol. Method Theor 26, 2019, 943-966; M.E. Smith, Energized crowding and the generative role of settlement aggregation and urbanization, in: A. Gyucha (Ed.), Coming Together: Comparative Approaches to Population Aggregation and Early Urbanization (Albany 2019), 37-58.
[8] E. R. Crema, A simulation model of fission–fusion dynamics and long-term settlement change, *Journal of Archaeological Method and Theory 21*, 2014, 385-404; A. Palmisano, Confronting scales of settlement hierarchy in state level-societies: Upper Mesopotamia and Central Anatolia in the Middle Bronze Age, JASc: Reorts 14, 2017, 220-240.
[9] T.A. Kohler - S. Cole - S. Ciupe, Population and warfare: a test of the Turchin model in Puebloan societies. in S. Shennan (Ed.), Pattern and Process in Cultural Evolution (Berkeley 2009), 277-295; P. Turchin - S.A. Nefedov, Secular Cycles (Princeton 2009).

with the introduction of farming could have caused a substantial population increase[10]. Such discussions feed the ongoing debate if technological and agricultural innovations developed in response to dramatic population growth or vice versa[11]. However, the rise and fall of the population can be attributed to a wide range of factors that may have interplayed with each other to some degree such as abrupt climatic shifts, warfare, resource depletion, epidemics, shifts in subsistence strategies and/or sociopolitical organisation, and large-scale migrations.

Bearing these issues in mind, modelling past human population trends is crucial for building long-term narratives explaining cultural and environmental change and the evolution of socio-ecological systems. How can we reconstruct past populations? The first challenge in archaeology is to identify those archaeological materials that might provide the most reliable indirect measures of population (see Drennan et al. 2015 for a good overview)[12]. Population estimates build on the assumption that an observable density of archaeological evidence over time and across a study region is somehow proportional to population. Put simply, the bigger the population, the stronger the signal in the archaeological record (e.g., the higher the density of pottery sherds, stone tools, site counts, radiocarbon dates, houses, etc.). While such indicators offer an idea of *relative* intensities of population and proportional change through time, they are more problematic for estimating *absolute* numbers of people in the past. Therefore, the first step for modelling past human population trends from a long-term perspective is to identify those archaeological proxies that may represent reliable indirect evidence of human activity and that are the least prone to any research biases and errors. In the past decades, the most popular archaeological proxies for modelling demographic trends have been 1) the raw count of archaeological sites, 2) the estimated extent of settlements, 3) and the summed probability distributions (SPD) of calibrated radiocarbon dates. Nevertheless, these proxies have been rarely compared directly. Thus, in this chapter, I advocate the use of a multi-proxy approach as a way forward in archaeological demography to have a better and more complete understanding of population dynamics in a given region during a certain chronological period. In practical terms, the divergences and convergences among the patterns defined by each archaeological proxy provide powerful insights and a wider range of explanations in describing population fluctuations both in the time span as a whole and in particular sub-periods.

The chapter of this book has been designed to provide a basic general introduction and toolkit for those interested in inferring past population dynamics from archaeological proxies. I first begin with a review of the three archaeological proxies mentioned above and the issues related to any of them. I then introduce and explain in detail the statistical methods for converting those archaeological population proxies into past demographic trends. Finally, conclusions are drawn with regard to the strengths and weaknesses of each method and archaeological proxy and their potential for understanding cycles of booms and busts in human population.

2. From radiocarbon dates to summed probability distribution (SPD)

The last few years have seen a dramatic increase of studies using large lists of archaeological radiocarbon (^{14}C) dates as a proxy for inferring time series of demographic trends in prehistory. The pioneering work by Rick lied the foundations of the so-called "dates as data approach", which relies on the assumption that the more people, the more anthropogenic products in the

[10] F. Silva – M. Vander Linden, Amplitude of travelling front as inferred from 14C predicts levels of genetic admixture among European early farmers, Scientific Reports 7(1), 2017; S. Shennan, The first farmers of Europe (Cambridge 2018), 110-114; N. Roberts – J. Woodbridge – A. Bevan – A. Palmisano – S. Shennan – E. Asouti, Human responses and non-responses to climatic variations during the last glacial-interglacial transition in the eastern Mediterranean, Quaternary Science Reviews 184, 2018, 47-67.

[11] E. Boserup, The Condition of Agricultural Growth. The Economics of Agrarian Change under Population Pressure (London 1965); M.N. Cohen, Food Crisis in Prehistory: Overpopulation and the Origins of Agriculture (Yale 1977).

[12] R. D. Drennan – C. A. Berrey – C. E. Peterson, Regional Settlement Demography in Archaeology (New York 2015).

archaeological record, the more organic samples to be recovered and then dated by radiocarbon dating resulting in a larger list of ^{14}C dates published[13]. The original approach consisted of generating histograms of uncalibrated radiocarbon ages. Then, from the early 1990s, some authors switched from histograms of uncalibrated radiocarbon ages to curves of calibrated radiocarbon dates[14]. The construction of summed probability distributions involves two steps: 1) the radiocarbon age of each sample (represented in the form of a Gaussian distribution on the y-axis, see Fig. 1) is transformed, using the calibration curve, into a probability distribution representing calendar years (represented in the x-axis, see Fig. 1); 2) the probability distributions of each sample are summed. Hence, the resulting curve is to be considered as a measure of dates intensity per calendar year (see SPDs in Figure 2).

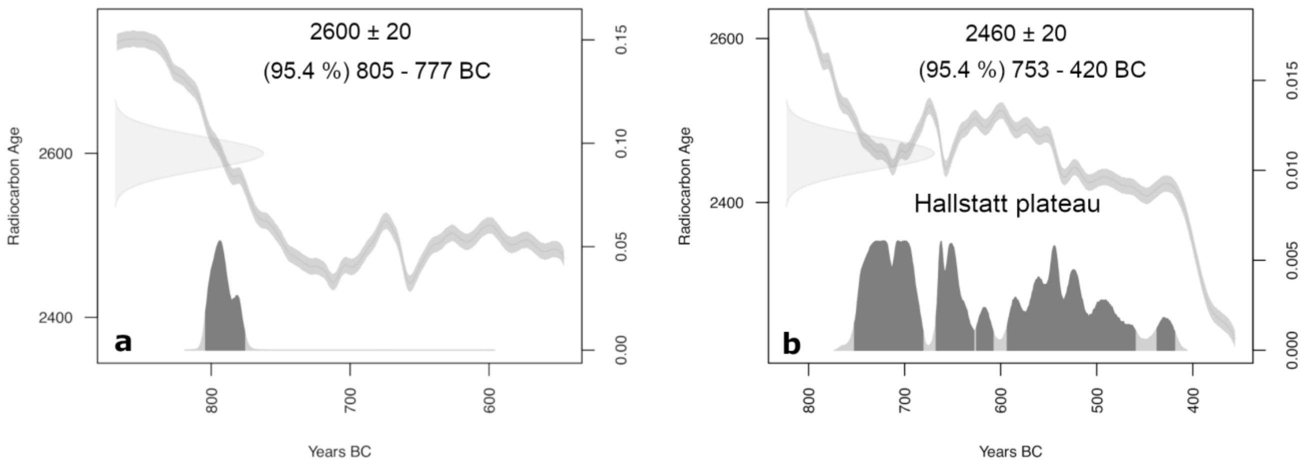

Figure 1. Graphs showing a single date at a) a steep portion of the calibration curve and at b) a flat portion of the calibration curve. Both graphs show the uncalibrated radiocarbon age and the calibrated radiocarbon date with a two-standard deviation confidence range (95.4%) in calendric years.

However, when dealing with calibrated radiocarbon dates is important to take in mind that the resulting SPD depends on the shape of the calibration curve. For instance, ^{14}C dates within steep portions of the calibration curve will produce narrower and spiky probability distributions of calibrated radiocarbon dates (see Fig.1a), while ^{14}C dates within flat portions of the calibration curve will result in wide and flat probability distributions (see Fig. 1b). This occurs even if the organic samples were dated at high levels of precision with an associated measurement error of ±15-20 years. Weninger and colleagues were the first ones to notice that normalised calibrated dates produce abrupt, artificial peaks in SPDs at steep portions of the radiocarbon calibration curve and opted for unnormalised dates to mitigate such an issue[15]. Figure 2a shows that the resulting SPD with normalised dates is characterised by pronounced spikes that are absent in the SPD generated by summing probability distribution from unnormalised calibrated radiocarbon dates (Fig. 2b). In addition, the period affected by the Hallstatt radiocarbon calibration plateau (ca. 800–400 BC) has discouraged most researchers from collecting radiocarbon samples belonging to archaeological contexts occurring within this time span (see Fig. 1b). As a consequence, uncalibrated radiocarbon age of around 2450 will always calibrate to ca. 800-400 BC, even in case of small measurement errors[16]. In fact, while the calibrated radiocarbon date in Fig.1a will have a 95.4 % of chance that the original organic samples will lie within a short time

[13] J. W. Rick, Dates as data: an examination of the Peruvian Preceramic radiocarbon record, American Antiquity 52, 1987, 55–73.
[14] K. M. Ames, The archaeology of the Longue Durée: Temporal and spatial scale in the evolution of social complexity in southern northwest coast, Antiquity 65, 1991, 935–945; J. C. Chatters, Population growth, climatic cooling, and the development of collector strategies on the southern plateau, Western North America, Journal of World Prehistory 9, 1995, 341–400.
[15] B. Weninger – L. Clare – O. Jüris – R. Jung – K. Edinborough, Quantum theory of radiocarbon calibration, WorldA 47, 2015, 543–566.
[16] J. Van der Plicht, Radiocarbon, the calibration curve and Scythian chronology, in E. M. Scott – A. Y. Alekseev – G. Zaitseva (eds.), Impact of the Environment on Human Migration in Eurasia, 2005, 45-61

span between 805 and 777 BC, the calibrated radiocarbon date in Fig. 1b lies within a much longer time span (753 – 420 BC) because the plateau in the calibration curve.

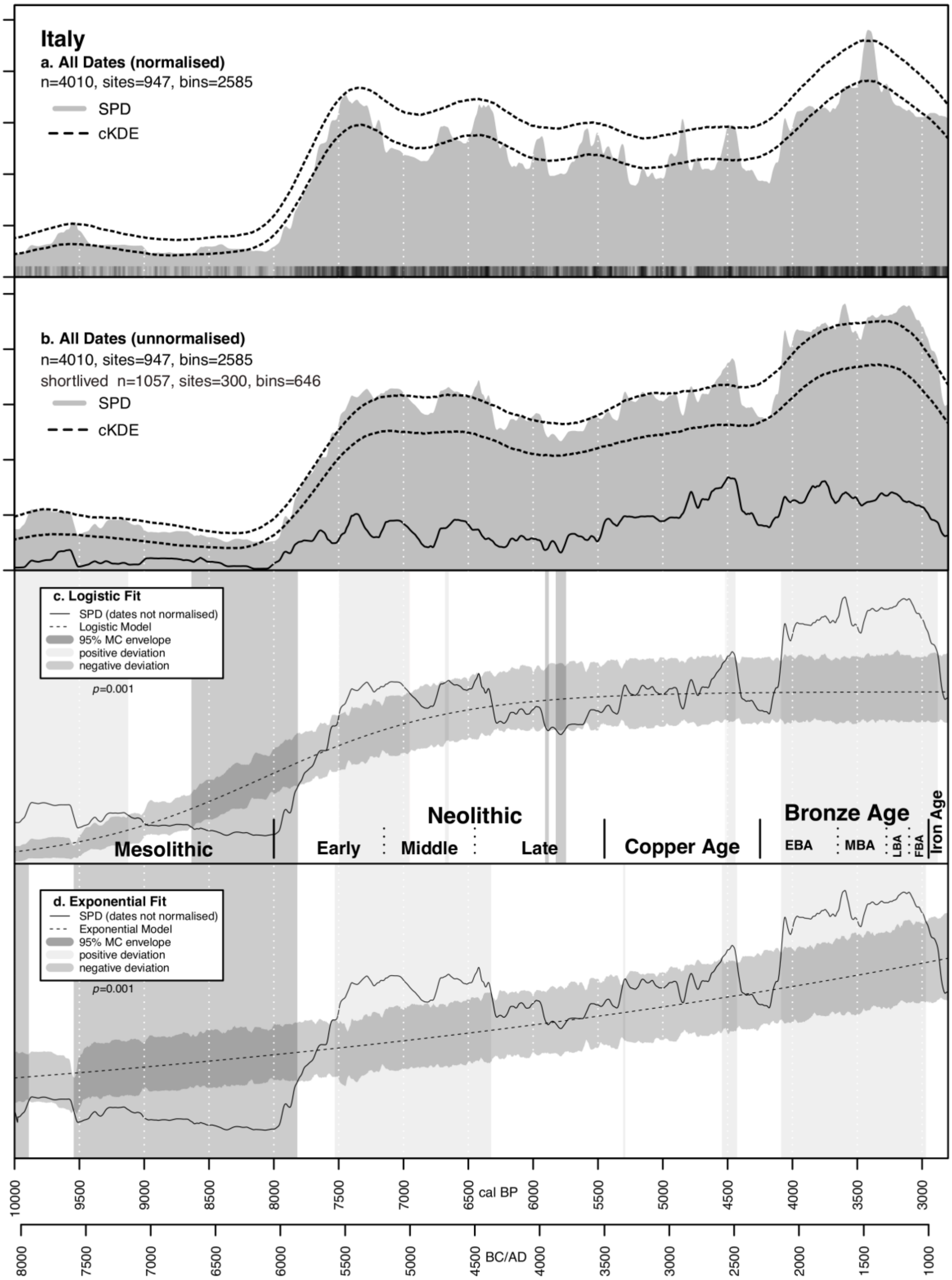

Figure 2. Summed Probability Distribution (SPD) of **a** normalised and **b** unnormalised calibrated radiocarbon dates; **c** unnormalised (solid line) vs. a fitted logistic; **d** exponential null model (95 % confidence grey envelope). Dark grey and light grey vertical bands indicate respectively chronological ranges within the observed SPD which deviate negatively and positively from the null models. The bar-code-like strip **a** represents the median values of multiple calibrated radiocarbon bins. Bootstrapped

composite Kernel Density Estimation (cKDE) of **a** normalised and **b** unnormalised calibrated radiocarbon dates are shown as an envelope bounded by dashed lines (data from Palmisano et al. 2021[17]).

In addition to the issues highlighted above, it is worth reviewing a series of other potential issues associated with the radiocarbon dates: 1) sampling error, 2) heterogeneity in sampling strategy and intensity, 3) taphonomic loss, and (4) old wood effect.
As said at the beginning of this chapter, population estimates build on the assumption that the number of archaeological evidence (e.g., sites, pottery sherds, radiocarbon dates) is directly proportional to the number of people. However, the relationship between the human population and the frequency of radiocarbon dates is not necessarily straightforward and can be affected by the actual number of collected radiocarbon dates for a given study area. In other words, the peaks and throughs of the SPDs of calibrated radiocarbon dates may not depict genuine population dynamics in the past. Undoubtedly a larger sample size may mitigate the issue of sampling error and some scholars, on the basis of former simulation-based studies, have suggested a minimum threshold of 500 dates to produce a reliable SPD for a time interval of around 10,000 years[18]. Although a "magical" threshold can be a reassuring starting point, a reliable sample size depends on several factors such as the extent of the area under scrutiny, the chronological scope of the investigation and the magnitude of the demographic phenomena to be analysed.

One problem arising when dealing with radiocarbon dates is that the available radiocarbon dates in a given study area do not often meet the assumption of complete randomness of statistical samples. Put simply, the radiocarbon dates used for inferring past population dynamics in a specific case study are the result of aggregating lists of published uncalibrated radiocarbon dates related to former studies applying diverse sampling strategies for a set of different purposes. Heterogenous sampling intensity across time and space is the most problematic issue undermining the reliability of the "dates as data approach" and could be related to several causes such as research biases, budget, the visibility of anthropogenic artefacts, methods, etc. In this context, certain chronological periods are more likely to be sampled than others. As a consequence, the interest in dating the introduction of farming or other specific periods (e.g., the transition from Late Pleistocene – Early Holocene, the subdivision of culturally defined periods such as the Bronze Age, etc.) could lead the diggers of the archaeological sites to collect organic samples from specific layers to the expense of others. For instance, the dramatic peaks visible in the modelled SPDs in the early stages of the Neolithic period could be the result of a higher-intensity sampling rather than a genuine pattern in population dynamics. By contrast, for the later historical periods, there is typically more reliance on cheaper dating methods based on documents, coins, and fine-ware pottery and less interest in paying for expensive radiometric dating. It follows that the radiocarbon dates are less reliable as a population proxy for the recent periods given the heterogenous sampling intensity which varies geographically from context to context and undermines the possibility of cross-regional studies. In fact, the recent cross-comparative work carried out under the umbrella of the Leverhulme-Trust funded project "Changing the Face of the Mediterranean" has suggested that radiocarbon dates can be a quite reliable proxy for reconstructing demographic trends up to ca. 800 BC[19] in the Mediterranean basin. For periods later than 800 BC the use of radiocarbon dates would be problematic and

[17] A. Palmisano – A. Bevan – A. Kabelindde – N. Roberts – S. Shennan, Long-Term Demographic Trends in Prehistoric Italy: Climate Impacts and Regionalised Socio-Ecological Trajectories, Journal of World Prehistory, 2021, 381–432.
[18] D. J. Michczynska – A. Pazdur, Shape analysis of cumulative probability density function of radiocarbon dates set in the study of climate change in late glacial and Holocene, Radiocarbon 46 (2), 2004, 733-744; A. N. Williams, The use of summed radiocarbon probability distributions in archaeology: A review of methods, JASc 39, 2012, 578–589.
[19] A. Bevan – A. Palmisano – J. Woodbridge – R. Fyfe – C. N. Roberts – S. Shennan, The changing face of the Mediterranean – Land cover, demography and environmental change: Introduction and overview, The Holocene 29(5), 2019, 703–707; C.N. Roberts – J. Woodbridge – A. Palmisano – A. Bevan – R. Fyfe -S. Shennan, Mediterranean landscape change during the Holocene: synthesis, comparison and regional trends in population, land cover and climate, Holocene 29 (5), 2019, 923-937.

would result in an underestimation of the real population. Further issues affecting the intra-site recovery of organic samples are the research budget and methods employed by the archaeological teams. For example, university-led research projects funded by large grants are more likely to employ more advanced field techniques and radiocarbon-based methods to date archaeological layers. Figure 3a shows a typical example of heterogonous spatial sampling intensity across Italy due to research biases focussing on specific topics (e.g., the introduction of farming, the spread of metallurgy, the rise of early complex societies) and the regional discrepancy of budgets available for research teams and commercial archaeological units.

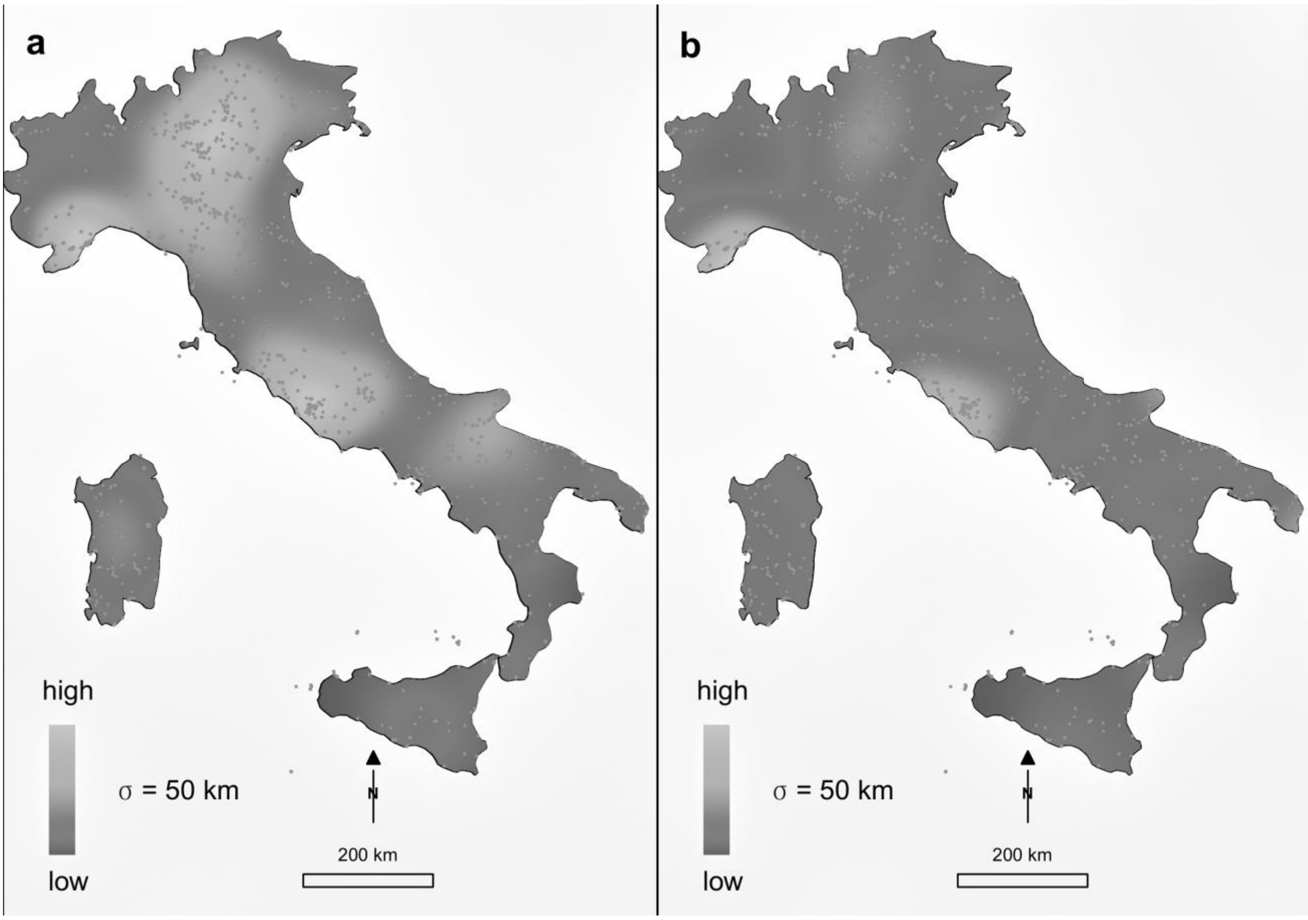

Figure 3. Kernel smoothed intensity of **a** un-binned and **b** binned radiocarbon dates (bandwidth = 50 km) from Italy as a five-category quantile classification (data from Palmisano et al. 2021).

The taphonomic loss and other natural and cultural post-depositional processes (e.g., agriculture, erosion, alluviation and colluviation, bioturbation, human excavations, wind deflation, etc.) are other key factors affecting the visibility and density of the archaeological record in a given region. Several studies have highlighted a greater rate of taphonomic loss for earlier archaeological deposits that are underrepresented when compared with more recent deposits[20]. Bluhm and Surovell have recently proposed a method to adjust the frequency of data based on the impact of taphonomic processes on the archaeological record[21].

The so-called old wood effect represents a pitfall when using radiometric dating because provides misleading and confusing results when materials belonging to different ages are deposited in the same archaeological context. For example, the radiocarbon dating of charcoal is

[20] T.A. Surovell – P. J. Brantingham, A note on the use of temporal frequency distributions in studies of prehistoric demography, JASc 34, 2007, 1868-1877; T. A. Surovell – J. Byrd Finley – G. M. Smith – P. J. Brantingham – R. Kelly, Correctingtemporal frequency distributions for taphonomic bias, JASc 36, 2009, 1715-1724.

[21] L. E. Bluhm – T. A. Surovell, Validation of a global model of taphonomic bias using geologic radiocarbon ages, Quaternary Research 91, 2019, 325–328.

prone to two potential tricky scenarios: 1) the charcoal can belong to roof timbers several centuries old when destroyed by fire, so one is dating the earlier period of construction of a building rather than its destruction; 2) some older wood beams could be reused in later contexts. For this reason, short-life organic samples such as bones, twigs and cereal grains are preferred. Therefore, when reconstructing SPDs of calibrated radiocarbon dates is worth comparing the SPD resulting from all available dates (grey area in Fig. 2b) with the one generated by using only radiocarbon dates from short-lived organic samples (black solid line in Fig. 2b). In our example, Figure 2b shows that the SPDs with all radiocarbon dates and with only short-lived dates are highly correlated ($r = 0.92$, p-value < 0.01, Pearson) and that there is not a significant old wood effect in the resulting post-calibration probability densities.

2.1 Addressing biases and statistical testing

This brief review of any potential biases when dealing with radiocarbon dates for inferring past demographic trends reminds us that the patterns depicted by SPDs are to be considered cautiously. Hence, any potential bias hidden in our dataset should be carefully assessed to mitigate any possible misinterpretation of patterns that could be statistical artefacts. In this section, I will illustrate a range of novel approaches that in the past decade contributed to diminishing the impact of these issues. The ubiquity of digital datasets of radiocarbon dates freely available and an increasing emphasis in archaeology to make all analyses reproducible represent a solid starting point for those who, for the first time, are willing to learn methods for modelling population dynamics by using radiocarbon dates[22].

Nowadays, the readers interested in applying specific techniques for inferring population dynamics from radiocarbon dates can benefit from the extension package *rcarbon* written for R[23], which provides basic tools for the calibration, aggregation of radiocarbon dates and the visualization and statistical inference of SPDs[24]. Another very useful R package is *c14bazAAR*, which allows for the querying, managing and merging of various open-access radiocarbon archives[25].

The first step when dealing with uncalibrated radiocarbon dates is to reduce any biases due to heterogeneity in sampling strategy and intensity. As said above, a particular well-funded research project or specific research targets could result in the oversampling of specific sites or intra-site site phases. In the hypothetical scenario shown in Figure 4, the layer (or site) A has yielded 7 radiocarbon dates while the equal size layer (or site) B has yielded only 2 radiocarbon dates. This problem can be mitigated by creating artificial binning, which consists in aggregating uncalibrated radiocarbon dates from the same layer (or site) that are within a defined-user number of years of each other and dividing by the number of dates that fall in this bin[26]. In the

[22] B. Marwick, Computational reproducibility in archaeological research: basic principles and a case study of their implementation, Journal of Archaeological Method and Theory 24, 2017, 424–450; K. Manning – S. Colledge – E. R. Crema – S. Shennan – A. Timpson, The cultural evolution of Neolithic Europe. EUROEVOL Dataset 1: Sites, phases and radiocarbon data, Journal of Open Archaeology Data 5, 2016 (2); G. Lucarini – T. Wilkinson – E. R. Crema – A. Palombini – A. Bevan – C. Broodbank, The MedAfriCarbon radiocarbon database and web application. Archaeological Dynamics in Mediterranean Africa, ca. 9600–700 BC, Journal of Open Archaeology Data 8 (1), 2020; A. Palmisano – A. Bevan – D. Lawrence – S. Shennan, The NERD Dataset: Near East Radiocarbon Dates between 15,000 and 1,500 cal. yr. BP, Journal of Open Archaeology Data 10(2), 1-9.

[23] R is a programming language for statistical computing which has become increasingly popular among social and natural scientists in the past decade; R Core Team 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

[24] For the techniques enabled in the R package *rcarbon* see E. R. Crema – A. Bevan, Inference from large sets of radiocarbon dates: software and methods, Radiocarbon 63(1),2021, 23-39; An R vignette associated with the package will drive step by step any new practitioners interested in applying this kind of analyses to their own dataset: https://cran.r-project.org/web/packages/rcarbon/vignettes/rcarbon.html.

[25] C. Schmid – D. Seidensticker – M. Hinz, c14bazAAR: An R package for downloading and preparing C14 dates from different source databases, Journal of Open Source Software 4(43), 2019, 1914.

[26] A. Timpson – S. Colledge – E. R. Crema – K. Edinborough – T. Kerig – K. Manning – M. G. Thomas – S. Shennan, Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: A new case study using an improved method, JASc 52, 2014, 549–557.

example shown in Figure 4, a 100-year bin produces two radiocarbon dates from site A and one date from site B. Once this step is done for all sites, the probabilities from each bin are summed. This procedure ensures that each site-phase is equally weighted and mitigates the spatial and chronological inhomogeneity of radiocarbon dates available for a specific study area (cf. Fig. 3b).

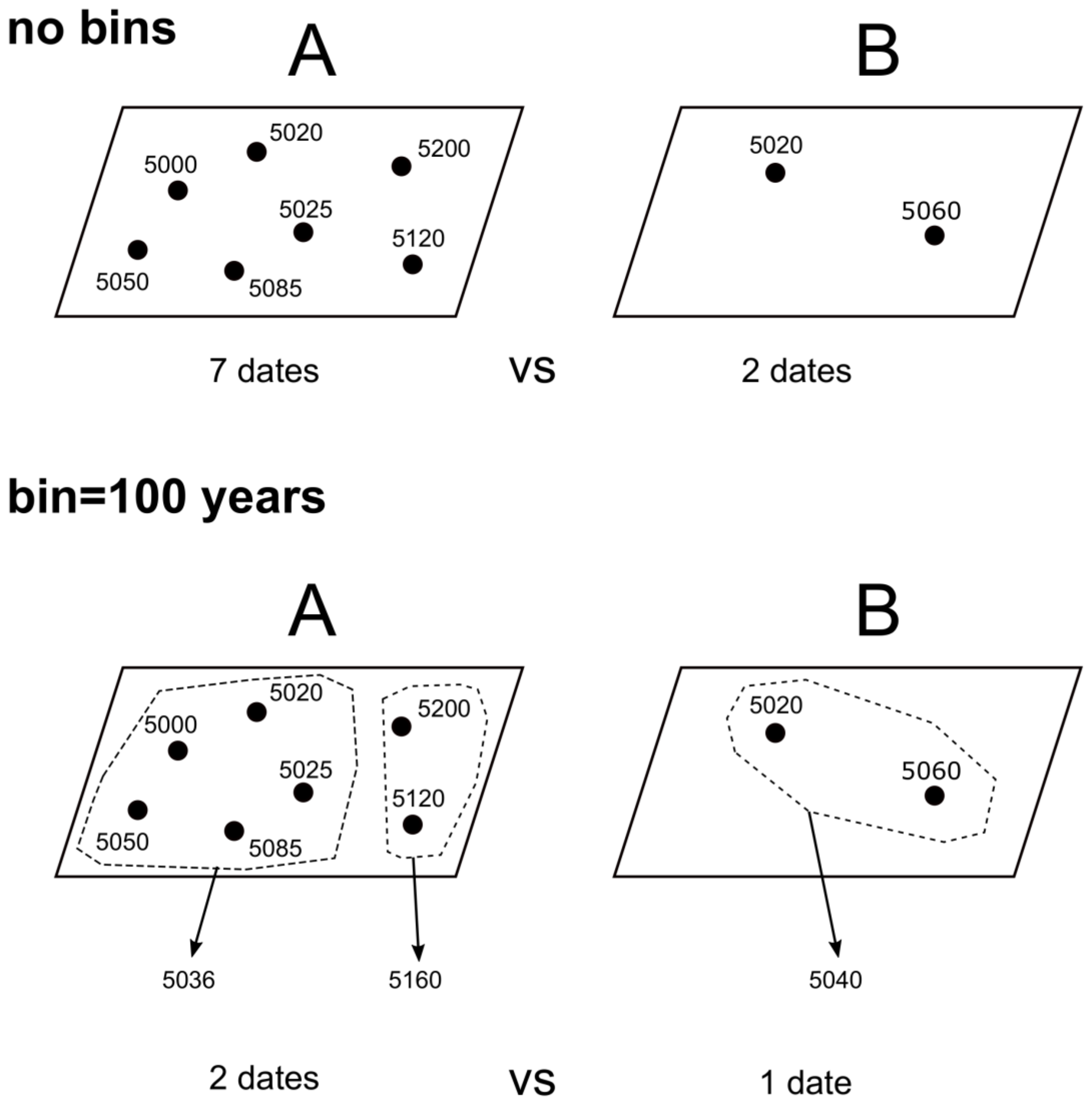

Figure 4. Different scenarios from layers (or sites) A and B with un-binned and binned radiocarbon dates.

The selection of an appropriate temporal threshold to aggregate dates into bins could produce different results and should be explored via sensitivity analyses. Figure 5 shows different SPDs by using different cut-off values for binning. Although the absolute summed probabilities change according to a specific temporal threshold, all SPDs depict similar patterns in terms of relative intensities of the population with dramatic peaks occurring at around 8000 cal. yr. BP and 4000 cal. yr. BP. Nevertheless, it is important to consider that this approach could underestimate the real population because it does not take into account the size of the site (dates from the same site and phase are lumped, whether the site is Rome or a small farmstead). In other words, this technique adopts a conservative approach which postulates that the frequency of radiocarbon dates for certain site-phases comes from biases in the intensity of investigation rather than the larger size of the site in that phase.

Given the multiple biases to which the radiocarbon dates are prone, the reconstructive method of generating SPDs from calibrated radiocarbon dates could be exposed to limitations and the

simple eyeballing of the reconstructed population fluctuations cannot be methodologically appropriate. Therefore, the seminal work by Shennan and colleagues introduced a Monte-Carlo based simulation approach to test statistically if the inferred demographic trends describe meaningful patterns not derived by mere chance[27]. This approach consists in comparing the observed SPDs against a theoretical null model of demographic change (e.g., uniform, exponential, logistic). This technique involves three steps: 1) a null model of demographic growth is fitted to the observed SPD, 2) random samples (equal to the number of bins) are drawn from the fitted model and then back-calibrated, 3) the resulting radiocarbon dates are calibrated and their probability distributions summed in order to generate an expected SPD of the fitted model. This process was repeated 1000 times to produce a 95 % confidence envelope (in grey in Fig. 2c-d). Deviations above and below the 95 % confidence limits of the envelope, respectively, indicate periods of population growth (in light grey) and decline (in dark grey) greater than expected according to a logistic (Fig. 2c) or exponential model of population growth (Fig. 2d). Because 5% of the observed SPD could fall outside the confidence interval by pure chance, a global p-value has also been calculated to assess the area of the
observed SPD outside the confidence envelope. It is worth pointing out that this value takes into account the overall shape of the SPD and, therefore, it is not unusual to have global p-values that are not statistically significant even when positive or negative local deviations
are detected. A p-value lower than 0.05 indicates that the observed SPD statistically depart from a null model of population growth across the whole chronological scope under analysis. Of course, we already know that an observed SPD would unlikely have an exponential or logistic growth rate but what matters is to statistically assess if the peaks and throughs depicted are significant and not mere statistical artefacts due to the underlying biases of the original dataset.

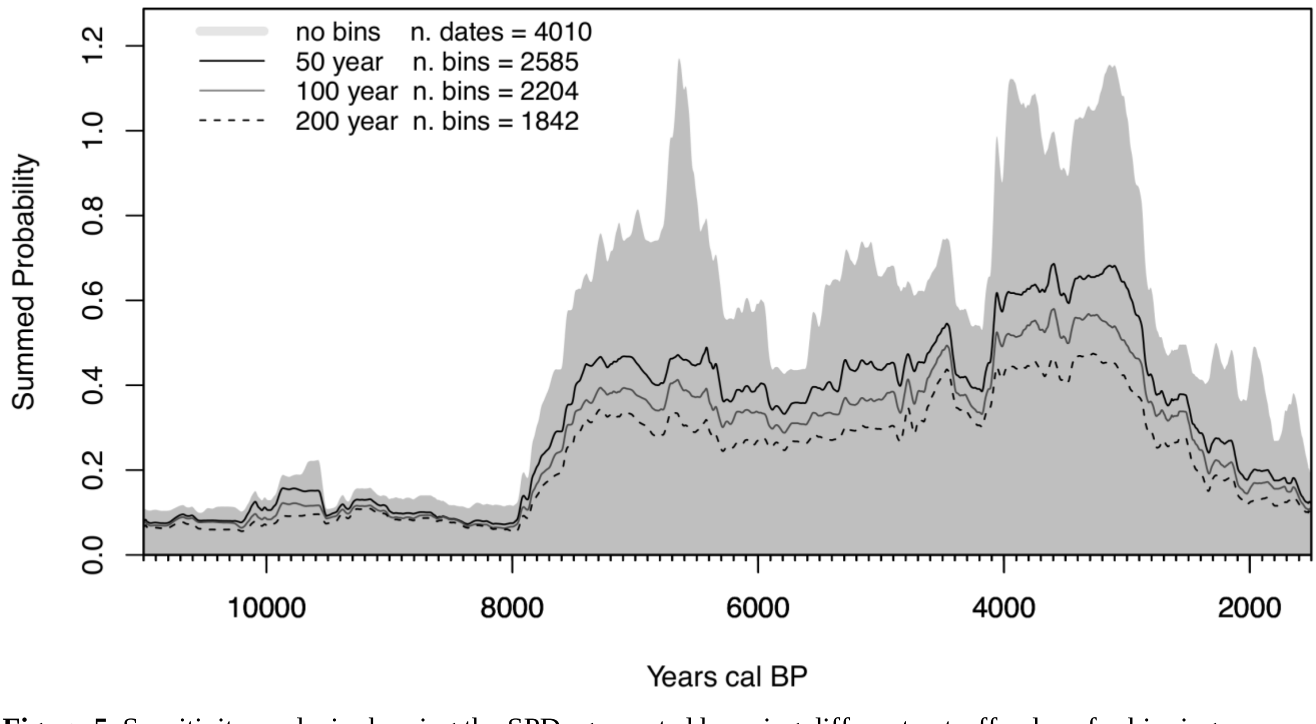

Figure 5. Sensitivity analysis showing the SPDs generated by using different cut-off values for binning.

An alternative approach to summed probability modelling is generating bootstrapped composite kernel density estimations (cKDE). First, time bins are randomly sampled with replacement, then a calendar date is randomly drawn from the probability density of each calibrated radiocarbon

[27] S. Shennan – S. S. Downey – A. Timpson – K. Edinborough – S. Colledge – T. Kerig – K. Manning – M. G. Thomas, Regional population collapse followed initial agriculture booms in mid-Holocene Europe, *Nature Communications4*, 2013, 2486.

bin and a Gaussian kernel density is estimated using a user-defined bandwidth size (in years)[28]. These two steps are repeated 1000 times to produce a 95 % confidence envelope of cKDE (see Fig. 2a-b). If the confidence interval is narrow, the observed pattern likely depicts a good picture of reality. This approach has the merit of smoothing the calibration noise of SPDs of calibrated radiocarbon dates and addressing several issues such as the heterogeneity in sampling strategy and intensity, chronological uncertainty and calibration artefacts[29].

3. Counting sites and summing the estimated size of settlements

The count of archaeological sites and the extent of settlements have a long pedigree as proxies for inferring past human population dynamics over the long-run[30]. These two proxies have in common with the radiocarbon dates two issues such as the heterogeneity in sampling intensity and the taphonomic loss. The different sampling strategies employed in a ground reconnaissance survey (e.g., random, stratified random, systematic, etc.) associated with research interests resulting in specific periods and areas being better investigated than others can undermine the reliability of the available dataset. In addition, the intensity of archaeological surveys assumed as the amount of effort expended on a given spatial unit is not homogenous and is generally inversely proportional to the size of the investigated area[31]. As a consequence, the number of sites can be biased by the intensity of the archaeological ground surveys carried out in a given region. Other factors affecting the archaeological surveys are 1) post-depositional processes (e.g., alluvium, colluvium, erosion, modern bulldozing activities) that may have damaged and buried the sites of a given region, 2) the visibility related to the kind of vegetation or the properties of sites (e.g., mounds, tells), 3) and the accessibility of the area to investigate related to political boundaries, geographical features or conflict zones.

Furthermore, the count of sites can be problematic and underestimate the level of the population when settlement systems show patterns of urban nucleation in the form of a growing concentration of population into fewer large centres[32]. The area of a site is therefore a useful and complementary addition to site counts and relies on the assumption that the number of inhabitants is somehow proportional to the spatial extent of a settlement. However, this correlation does not scale linearly and is not universally consistent across different regions of the world[33]. For example, large urban centres are usually densely packed and show a higher population density per hectare than medium-sized and small settlements.

A further and crucial issue when dealing with archaeological survey data is their intrinsic chronological uncertainty and the struggle to date the foundation, the duration and the abandonment of sites. Without the support of radiometric dating, the occupation period of archaeological sites is dated through typo-chronological schemes defined by archaeological cultures that in the case of long-lived pottery or surface materials alone can produce site phases

[28] C. Bronk Ramsey, Methods for summarizing radiocarbon datasets. Radiocarbon 59(6), 2017, 1809–1833; W. A. Brown, The past and future of growth rate estimation in demographic temporal frequency analysis: Biodemographic interpretability and the ascendance of dynamic growth models, JASc 80, 2017, 96–108.

[29] E. R. Crema, 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, 2022, 1398-1401.

[30] W.T. Sanders, The Cultural Ecology of the Teotihuacan Valley (University Park, Pa,1965); R.M. Adams, Land Behind Baghdad: a History of Settlement on the Diyala Plains, (Chicago 1965).

[31] S. Plog – F. Plog – W. Wait, Decision making in modern surveys, in M. B. Schiffer (Ed.), Advances in Archaeological Method and Theory (New York 1978), 384-421; J. F. Cherry, Frogs round the pond: perspectives on current archaeological survey projects in the Mediterranean region, in D.R. Keller – D. W. Rupp (Eds.), Archaeological Survey in the Mediterranean Region (Oxford 1983), 375-416.

[32] C. E. Peterson – R. D. Drennan, Patterned variation in regional trajectories of community growth, in M. Smith (Ed.), The Comparative Archaeology of Complex Societies (Cambridge 2011), 88-137; A. Palmisano, Confronting scales of settlement hierarchy in state-level societies: Upper Mesopotamia and Central Anatolia in the Middle Bronze Age, JASc: Reports 14, 2017, 220-240.

[33] R. D. Drennan – C. A. Berrey – C. E. Peterson, Regional Settlement Demography in Archaeology (New York 2015), 20-25.

spanning thousands of years. We are also left uncertain if a given site was occupied seasonally or permanently and if the sites of a study area were all contemporary.

3.1 Dealing with temporal uncertainty

This section will showcase some methods addressing temporal uncertainty in the archaeological record. When using the count and extent of archaeological sites, the first step is to decide what kinds of sites are to be part of the counting exercise. Given that the term site may also refer to temporary activity areas (e.g., seasonal campsites), industrial zones (mines), infrastructures (e.g., bridges, harbours) and cemeteries, for inferring past population trends is preferred to employ in the analysis only those sites identified as human dwelling places.

The number and size of sites can be summed by making use of user-defined time steps (e.g., 100 or 200 years). Moreover, this approach can be problematic as it does not take into account the chronological uncertainty of the archaeological dataset. The site phases (or occupation periods) of a certain site have different time spans according to the dating precision provided by the archaeological materials: longer time spans with higher uncertainty occur when the dating is based on artefacts with little diagnostic value (e.g., long-lived pottery type), while shorter time spans occur thanks to the high precision dating of some artefacts such as dateable coins, written sources and short-lived pottery type). For example, in the gazetteer of archaeological surveys reports, some sites are generally recorded as Neolithic and are assigned a time span of up to two-three thousand years because of the low diagnostic power of the surface material collected (e.g., potsherds, lithics). In this context, the seminal work by Crema and colleagues, who first introduced aoristic approaches in archaeology, paved the way for a series of studies addressing the issue of temporal uncertainty in the archaeological record[34]. The method builds on the assumption that the total probability of an archaeological event (site occupation phase in our case) within a given time span is 1, which indicates an absolute certainty that the site was in use in that time span. If we then divide by the length of the site's chronological range, we can represent the probability of existence for each temporal block (implicitly therefore adopting a default uniform assumption). For example, using time steps of 200 years, a site phase of an archaeological site from Northern Italy is broadly dated to the Bronze Age according to the existing typo-chronological schemes of the pottery collected. In this case, the site phase ranging from 2200 to 1200 BC has an aoristic weight of 0.20 for each time-step (2200–2000, 2000–1800, 1800–1600, 1600–1400, 1400 – 1200; see Fig. 6c). Instead, a site-phase with a shorter time-span ranging from 2200 to 1800 BC has an aoristic weight of 0.5 for each time-step and so on (2200–2000, 2000–1800; Fig. 6e). In the case of a Middle Bronze Age sites dated to a period ranging from 1600 to 1300 BC the aoristic approach involves two steps (see Fig. 6b): 1) the probability value for each year is 0.0033 (the total probability of 1 divided by 300); 2) the probability of each year is multiplied by the number of years falling within each time step (0.0033 x 200 = 0.667 in the time step 1600-1400; 0.0033 x 100 = 0.33 in the time step 1400-1200 BC). Having assigned such weights to each site, we can then sum them all to obtain the aoristic sum for each temporal block (Fig. 6).

[34] E. R. Crema – A. Bevan – M. Lake, A probabilistic framework for assessing spatio-temporal point patterns in the archaeological record, JASc 37, 2010, 1118-1130; E.R. Crema Modelling temporal uncertainty in archaeological analysis, J. Archaeol. Method Theory 19 (3), 2012, 440-461; J. Kolář – M. Macek – P. Tkáč – P. Szabó, Spatio-temporal modelling as a way to reconstruct patterns of past human activities, Archaeometry 58 (3), 2016, 513-528; A. Palmisano – A. Bevan – S. Shennan, Comparing archaeological proxies for long-term population patterns: An example from central Italy, JASc 87, 2017, 59-72.

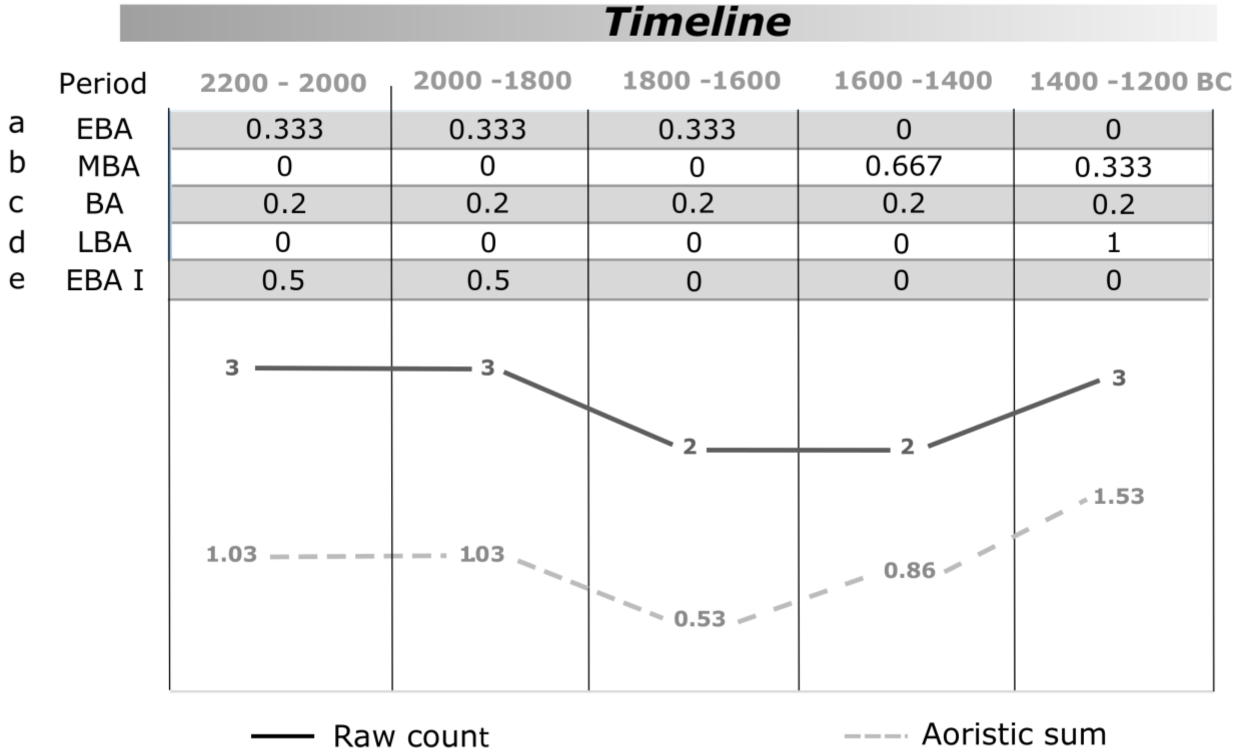

Figure 6. Five different site-phases (a-d) with varying chronological ranges indicated horizontally by the length of their temporal blocks (EBA: Early Bronze Age (2200-1600 BC), MBA: Middle Bronze Age (1600-1300 BC), BA: Bronze Age (2200-1200 BC), (LBA: Late Bronze Age (1300-1200 BC), EBA I: Early Bronze Age I (2200-1800 BC)). Each 200-year time-step reports an aoristic weight, which represents the probability of the existence of a site within each temporal block. The aoristic sum (dashed line) is plotted vs. the raw count of sites (solid line).

Aoristic weights change when the temporal resolution of the analysis is modified (aoristic weights would be lower with timesteps of 100 years). Furthermore, given that the aoristic analysis assumes a uniform probability distribution (as noted above), each year has the same probability of being the one in which the archaeological event occurred. Therefore, an issue arises when a large sample of archaeological events with the same period occurs within the same temporal block. This would flatten the curve depicting the demographic trend because of the similar temporal structure of a large number of sites beginning and ending at the same time. While it is often difficult to assess the likely longevity of an individual site without considerable amounts of absolute dating, it is often evident that site durations are shorter than their assigned chronological ranges. To mitigate this tension between wide chronological uncertainties and narrower likely site durations, the Monte Carlo simulation approach generates randomised occupation start dates for sites with low-resolution information by drawing a date from a uniform distribution within the chronological range of the site phase under question (see Fig. 7a in simulations 1-2). Then, a site duration randomly generated from a normal distribution with a mean of, for example, 200 years and a standard deviation of 50 years is added to the start date drawn (Fig. 7b in simulations 1-2). This typical duration is chosen to correspond to the modal site phase lengths from the dataset used in the analyse and offer clear contrast for those periods where uncertainties are much larger (e.g., thousands of years), but the choice of mean expected site duration is slightly arbitrary and would best be informed by a wider range of evidence. In any case, this approach allows one to deal with those site phases having a coarser resolution and periods ranging over thousands of years. We can then simulate multiple time series and generate a 95% critical envelope for all randomised start dates and durations of site occupation phases (see the grey envelope in Fig. 7). The width of the envelope is indicative of the degree of temporal uncertainty in site occupation through time. In other words, the narrower and larger portions of the Monte Carlo simulated envelope respectively indicate periods of lower and bigger uncertainty in the demographic trends depicted.

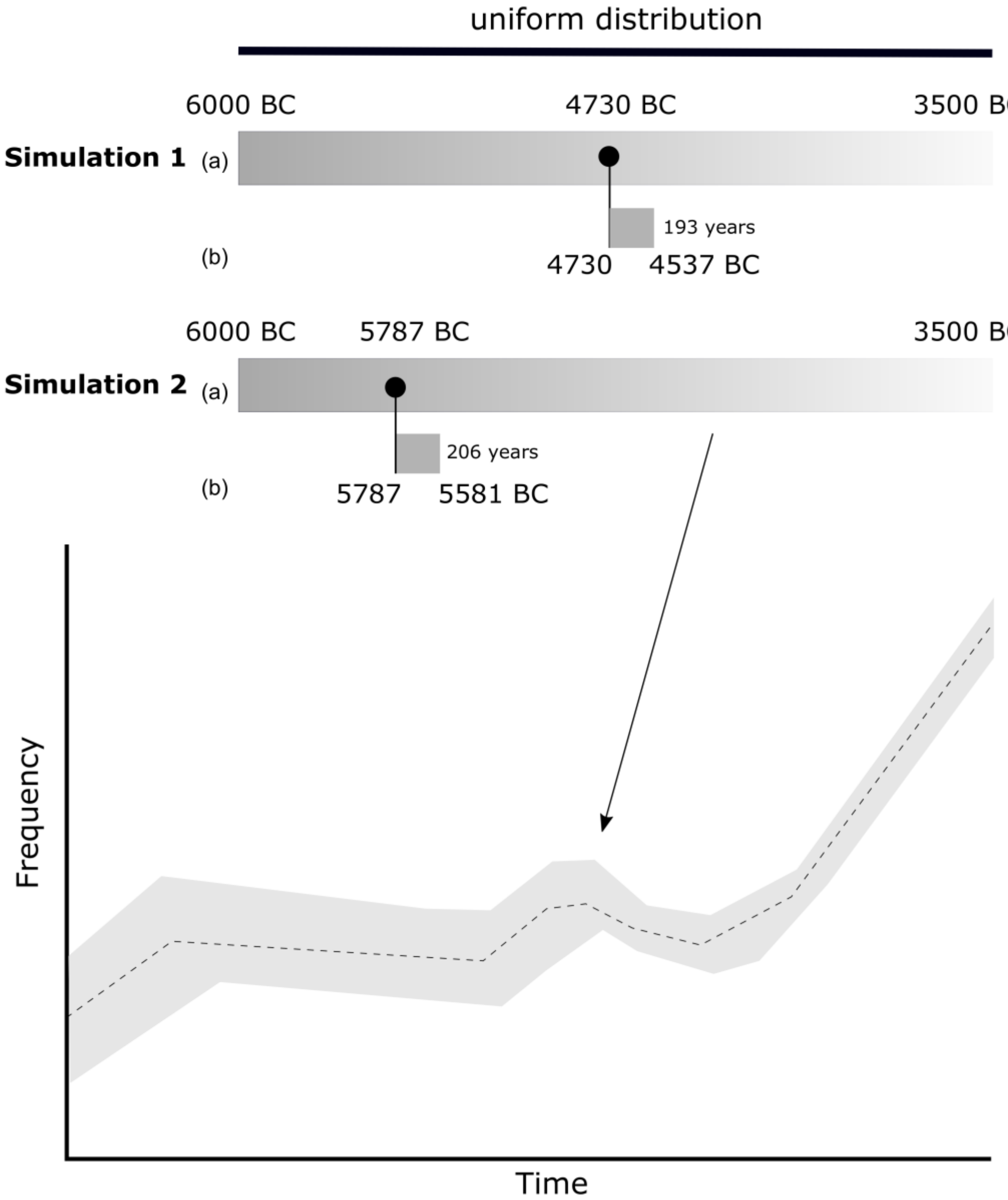

Figure 7. Randomised start date of sites: a) a date is randomly drawn within the time span of a given site phase with a uniform probability distribution and b) a site duration is randomly generated from a normal distribution with a mean of 200 years and a standard deviation of 50 years is added to the drawn date. In grey is represented the envelope of 1000 Monte Carlo simulated time-series of the randomised start date of sites.

4. Comparing multiple archaeological proxies

The survey above has provided an overall picture of the issues to which the three archaeological proxies here discussed are prone. However, I have here showcased some methods developed in the past decade to mitigate those issues. Unfortunately, radiocarbon dates, site count and estimated size of settlements are three proxies rarely compared directly[35]. Therefore, in this

[35] Here some examples of studies applying a multi-proxy approach: M. Tallavaara – P. Pesonen – M. Oinonen, Prehistoric population history in eastern Fennoscandia, JASc 37 (2), 2010, 251-260; J. C. French, The demography of the Upper Palaeolithic hunter–gatherers of Southwestern France: a multi-proxy approach using archaeological data, J. Anthropol. Archaeol. 39, 2015, 193-209; A. Palmisano – A. Bevan – S. Shennan, Comparing archaeological proxies for long-term population patterns: An example from central Italy, JASc 87, 2017, 59-72.

chapter, I advocate a multi-proxy approach to compare various independent archaeological indices for building more robust narratives about demographic trends over the *longue durée.*

Scholars working on prehistoric periods have mainly privileged the use of radiocarbon dates as proxies for reconstructing past demographic trends. This is comprehensible given that radiocarbon dates provide a better chronological resolution above all in those cases when site phases are broadly dated through typo-chronological schemes based on material culture (e.g., Upper Paleolithic, Neolithic, Copper Age). Another aspect is represented by the almost ubiquitous availability of digital datasets freely accessible online through numerous repositories (e.g., *Zenodo*, *figshare*, *Harvard Dataverse*, etc.) and data papers. In recent years, *Journal of Open Archaeology Data* (JOAD) has published several digital archives of radiocarbon dates respecting specific standards in terms of open licenses permitting unrestricted access (e.g., CC0, CC-BY) (https://openarchaeologydata.metajnl.com/). In addition, the community of scholars involved in palaeodemographic studies has been experiencing in recent years an increasing emphasis on Open Science practices, which propelled the free dissemination of datasets, reproducible research (in the form of scripts for running the analyses) and novel methods (see above the mentioned R packages *rcarbon* and *c14bazAAR*).

Unlike the radiocarbon dates, the archaeological survey data such as the site count and the estimated size of sites are not so accessible online and do not benefit from established and popular methods freely accessible among the practitioners. This is due to a series of reasons: 1) the archaeological survey data are usually used for modelling population dynamics in historical periods by archaeologists that methodologically are not as familiar with statistical and quantitative methods as prehistoric archaeologists; 2) the archaeological survey data are usually published via traditional media such as survey reports in hard paper copies which makes time-consuming extracting and converting any data into digital formats; 3) the prevalence of data hoarding practices typical of scholars less familiar with Open Science approaches. Unlike the radiocarbon dates, there are very few digital archives of archaeological survey datasets freely accessible online[36].

Ideally, the use of multiple archaeological proxies is preferable for building more robust narratives about past population dynamics. As we have seen in the former sections above, both radiocarbon dates and archaeological survey data are prone to any kind of biases and have their weaknesses and strengths. For sure, the radiocarbon dates provide a good advantage in terms of higher levels of resolution, which makes them the privileged dataset for inferring population fluctuations in prehistory and for understanding past human-environment interactions. The latter point needs higher resolution demographic proxies to compare versus paleoenvironmental proxies (e.g., paleoclimatic proxies, pollen diagrams, etc.). Unlike the radiocarbon dates, the archaeological survey data allows us to have better control of any biases derived by a heterogeneous sampling intensity. Unlike the radiocarbon dates, the archaeological survey reports provide information about the spatial coverage, the sampling strategy and the intensity of the ground reconnaissance survey carried out in the field. This allows us to understand which regions and periods are under or overestimated in our study area and consequently to be more aware of possible artefacts in the trends depicted in the reconstructed population curves. Furthermore, the archaeological proxies discussed in this chapter can be good indicators of the relative and proportional change of population across time but they may be problematic for modelling absolute numbers of people. In simple words, stating that the population is "two times bigger" could mean 10,000 as opposed to 5000, or it could mean 10 as opposed to 5. Of course,

[36] Recently, some data papers have disseminated digital archives of archaeological survey data. Here some examples: K. A. Crawford – M. A. Vella, Cyprus Dataset: Settlements from 11000 BCE to 1878 CE, Journal of Open Archaeology Data 10 (7), 2022; P. Pajdla – F. Trampota, Neolithic Settlements in Central Europe: Data from the Project 'Lifestyle as an Unintentional Identity in the Neolithic', Journal of Open Archaeology Data 9 (13), 2021; A. Palmisano – A. Bevan – S. Shennan, Regional Demographic Trends and Settlement Patterns in Central Italy: Archaeological Sites and Radiocarbon Dates, Journal of Open Archaeology Data 6 (2), 2018.

there is a big difference in terms of magnitude in the two examples above and a relative population change cannot be fully informative if we are modelling demographic trends in response to abrupt shifts in environmental conditions such as climate change or the exhaustion of local natural resources. A more complete understanding of the human impact on the landscape would be possible in the case of absolute population estimates. Among the three proxies discussed here, only the estimated size of settlements can be converted into absolute numbers of people. Of course, although we have the real size of a settlement in a given period, it would be problematic to infer the number of inhabitants as there is no constant linear correlation between people and size. One solution would be to carry out sensitivity analyses by applying different numbers of people per hectare (e.g., 100, 200, 400, and so on) and to assess, for instance, under what scenario the carrying capacity of a specific area is exceeded. Therefore, this methodological exercise may be used as a heuristic device to assess under which circumstances a specific society can respond to endogenous (remarkable growth of population) or exogenous (e.g., climate change) stress factors.

Finally, the need for a multi-proxy approach will be here promoted by making use of a practical example. Figure 8 shows the demographic trends in central Italy from Late Mesolithic to Early Iron Age (7000 – 800 BC) inferred by using radiocarbon dates and data from archaeological surveys carried out in Latium and Tuscany (site count and estimated settlement size)[37]. The cross-comparison of these three demographic proxies is useful to detect further insights towards a better understating of the processes determining regional settlement demography and settlement patterns. The first step is to make all the proxies comparable by standardizing all of them on a range between 0 to 1. Overall, the archaeological proxies depict similar trends as also shown by the statistically significant pairwise Pearson's correlations (ranging between 0.53 and 0.98, see Fig. 9) among all proxies for the period spanning from 7000 to 800 BC (Fig. 8).

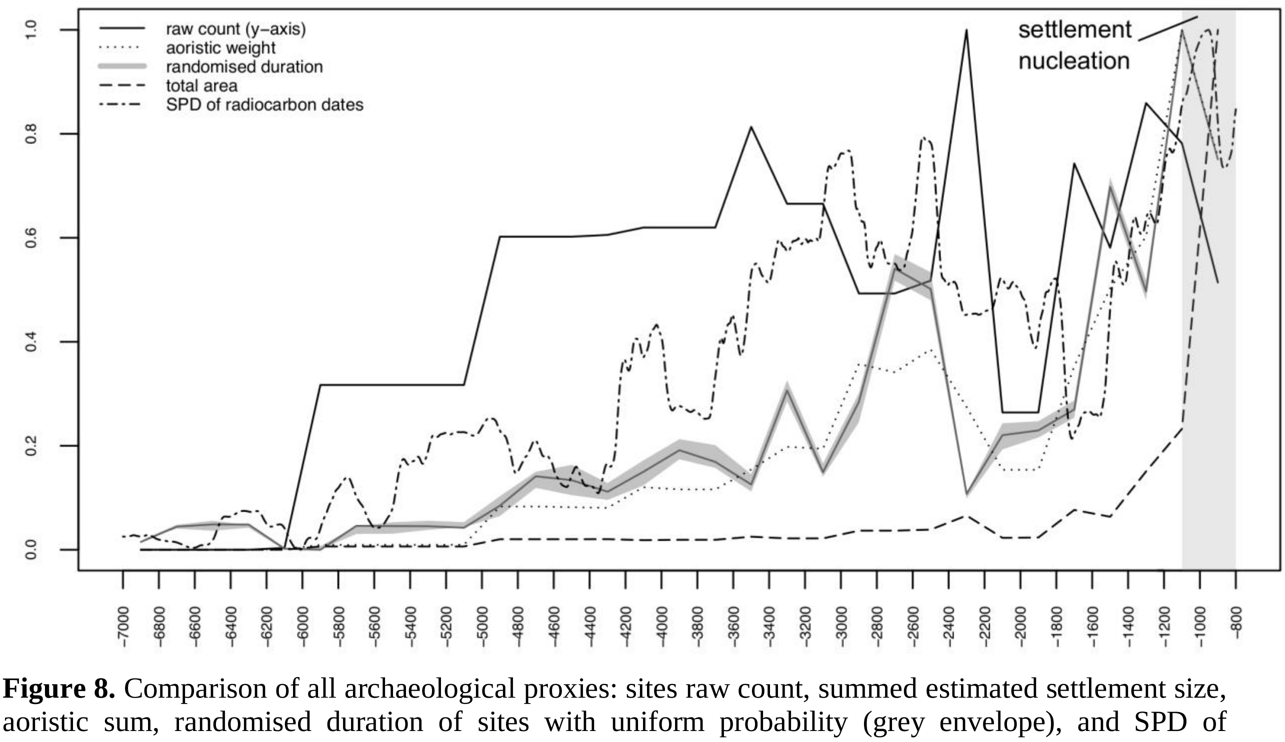

Figure 8. Comparison of all archaeological proxies: sites raw count, summed estimated settlement size, aoristic sum, randomised duration of sites with uniform probability (grey envelope), and SPD of radiocarbon dates from 7000 BC to 800 BC. All values have been standardised on a scale between 0 and 1 (Data from Palmisano et al. 2017[38]).

---

[37] The data come from the following data paper: A. Palmisano – A. Bevan – S. Shennan, Regional Demographic Trends and Settlement Patterns in Central Italy: Archaeological Sites and Radiocarbon Dates, Journal of Open Archaeology Data 6 (2), 2018.

[38] A. Palmisano – A. Bevan – S. Shennan, Comparing archaeological proxies for long-term population patterns: An example from central Italy, JASc 87, 2017, 59-72.

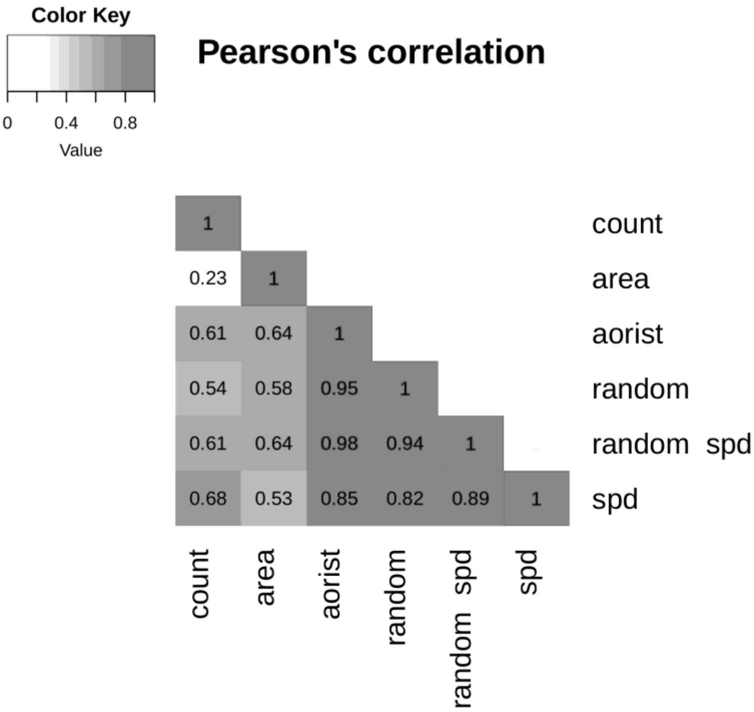

Figure 9. Pairwise Pearson's correlations between all archaeological proxies. The dark grey-white scale values represent correlation values, with dark grey representing the better fit and white the worse fit results.

All proxies show an increase in population in the early Neolithic (~ 6000 – 5500 BC), which seems to be associated with the earliest adoption of farming economies and a much more stationary population settled in permanent houses and villages[39]. A further substantial increase of population observable in all proxies occurs in the second half of the fourth millennium BC and results in two peaks during the Copper Age (3200-2800 and 2600-2400 BC) punctuated by a population decline between 2800 and 2600 BC (Fig. 8). In this period, communities started living in settlements greater than 1 ha in size and based on more intensive mixed economies (agriculture, hunting, herding) comprising specialized techniques in the cultivation of different kinds of cereals and pulses[40]. At a first glance, the second population peak occurring during the Copper Age at 2500 BC in the SPD seems to be shifted by 200 years from the one observed in the site count and the summed settlement size data (2300 BC, see Fig. 8). This could be the result of the difference in temporal resolution between the SPD of radiocarbon dates and the archaeological survey data, but regardless, it is likely that the peak at 2300 BC in the site count and summed estimated settlement size data is an artefact of temporal uncertainty in archaeological site-phases broadly dated to the Bronze Age and assigned to a broad period (2300 - 1000 BC). This is likely to overestimate the number of sites that effectively were occupied at the beginning of the Early Bronze Age. By contrast, the aoristic sum and the randomised site start dates do not report such a peak. After 1700 BC, all measures show gradual population growth, which peaked dramatically during the Final Bronze Age and the Early Iron Age (1200 - 900 BC). The divergence between the summed settlement size and the other settlement measures in this latter period is explained by the radical changes occurring in settlement patterns between the Final Bronze Age (ca. 1175-1020/950 BC) and the beginning of

---

[39] G. Barker, Prehistoric territories and economies in Central Italy, in E. Higgs (Ed.), Palaeoeconomy, (Cambridge 1975), 111-175; C. Malone, The Italian Neolithic: a synthesis of research, J. World Prehistory 17 (3), 2003, 267-269.

[40] A. P. Anzidei – G. Carboni, Il comprensorio dei Colli Albani ed aree limitrofe durante il Neolitico e l'Eneolitico, in L. Drago Troccoli (Ed.), Il Lazio dai Colli Albani ai Monti Lepini tra preistoria ed età moderna, (Roma 2009), 89-98;

the Early Iron Age (ca. 1020/950-750/25 BC) when nucleation occurred in the form of an extreme concentration of population into a few large centres. This period sees the abandonment of small-sized dispersed villages (generally 2-3ha) located in open positions or on small hilltops and the occupation of a smaller number of sites of larger sizes (50-100 ha) distributed over the lowlands and plateaus (see Fig. 10)[41]. Thus, during the Final Bronze Age, 287 sites measured a total size of 300 ha (Fig. 10), while in the Early Iron Age 212 sites measured a total size of 1200 ha (Fig. 10). The site counts and summed settlement sizes show no correlation (r = 0.23, Fig. 9), contrasting with a higher value (r = 0 .59) between those two proxies up to the Final Bronze Age (1175/1150-1020/950 BC). In other words, we need the summed site area information, as well as the various measures based on site counts to provide a valid picture. The main drawbacks of the archaeological settlements, as data, are their coarse chronological resolution in comparison with the radiocarbon dates.

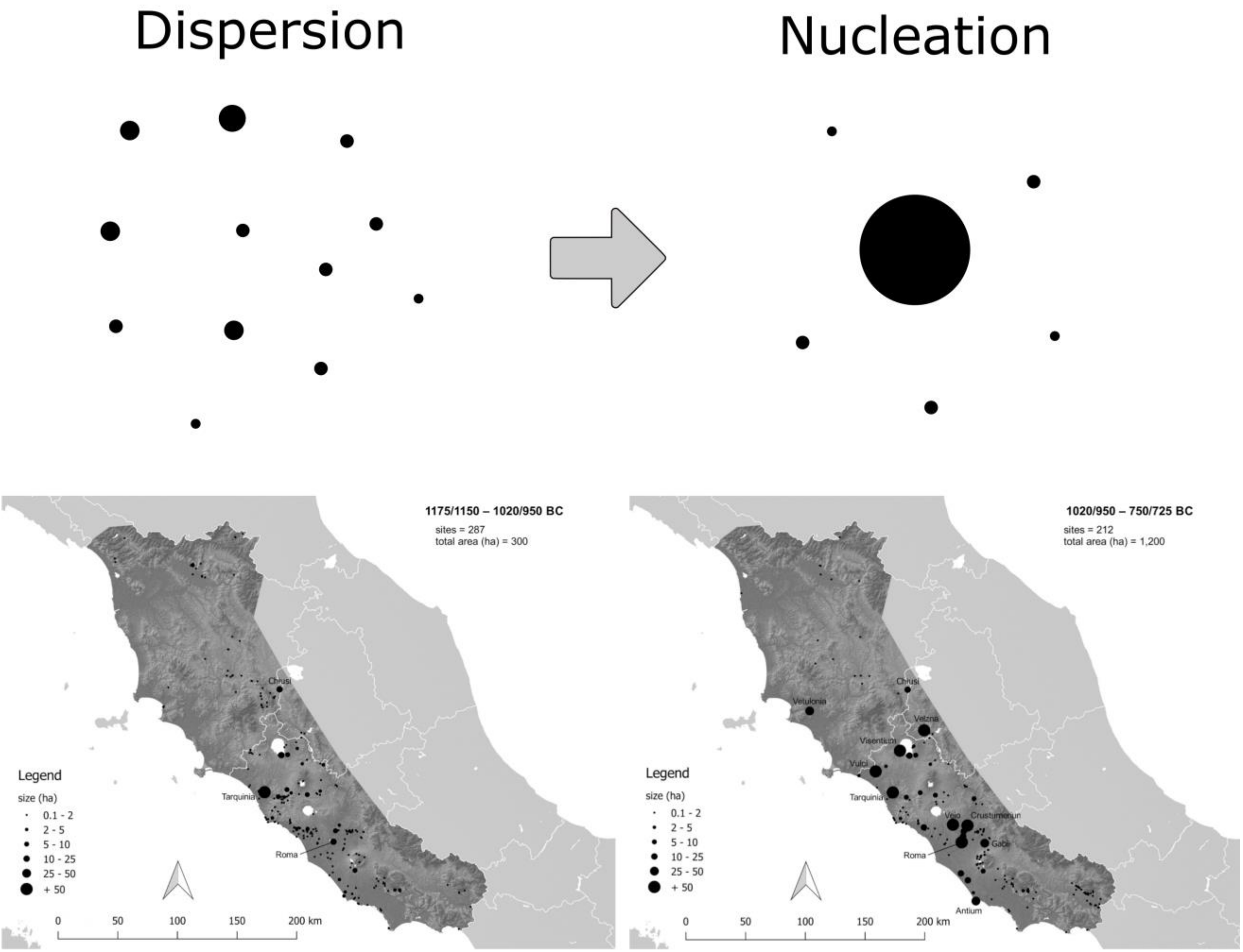

Figure 10. Patterns of settlements dispersion and nucleation respectively during the Final Bronze Age and the Early Iron Age.

5. Future perspectives and challenges

This chapter has provided an introduction and review of a wide range of well-established approaches commonly used for inferring past population dynamics from archaeological proxies. The methods discussed above have been developed for a variety of reasons addressing both the biases in the archaeological record and the need to assess quantitatively the significance of the

[41] R. Peroni, Formazione e sviluppo dei centri protourbani medio-tirrenici, in A. Carandini – R. Cappeili (Eds.), Roma, Romolo, Remo e la fondazione delta città, catalogo della mostra, (Milan 2000), 26-30; F. Fulminante, The Urbanisation of Rome and Latium Vetus: from the Bronze Age to the Archaic Era, (Cambridge 2014); L. Alessandri, Exploring territories: bubble model and minimum number of contemporary settlements. A case study from Etruria and Latium Vetus from the Early Bronze Age and Early Iron Age, Origini 37, 2015, 175-199.

trends detected. Some methods such as the binning function for the radiocarbon dates and the aoristic approach for the site count represent the foundation on which building specific analyses to address issues arising from the available archaeological datasets. Of course, there is no single optimal solution and users should ponder any option according to the quality of the original dataset and the goals of their research. I will here highlight a series of recommendations and simple guidelines that may be useful for novel users that just started approaching palaeodemographic approaches:

1) SPDs of calibrated radiocarbon dates and curves generated from archaeological survey data should be interpreted cautiously given the biases intrinsic in the original dataset. Of course, the methods shown above allow us to mitigate some of those biases but do not remove them at all. Therefore, the inferred population fluctuations are useful as a heuristic device, which provides important clues when dealing with a multi-millennial chronological scope. In particular, the approaches outputting envelopes defining the uncertainty of an event should be preferred in the arsenal of the available methods.

2) The number of radiocarbon dates and sites employed in the analyses should be appropriate according to the chronological scope and the spatial area under investigation. Of course, the larger the sample size the lower the effect of sampling error on a given dataset. However, a larger sample size represents just one side of the coin because in some cases we do not have a strong prior knowledge of the target population and the dataset used in our analyses could be not representative of the original sample.

3) Given that several methods shown above rely on some user-defined settings such as the temporal threshold for aggregating radiocarbon dates into bins and the bandwidth size in the Kernel Density Estimates, it is worth running sensitivity analyses. Although a user is able to justify the parameters chosen, any possible outcomes should be explored. The sensitivity analyses would reveal to what degree the reconstructed demographic curves change according to the settings. In the best-case scenario the resulting shapes of the curves will not differ, while in the worst-case scenario the results will depend on the parameters chosen. This kind of approach can be also useful when dealing with the reconstruction of the absolute number of people. In this latter case, sensitivity analyses applying a different number of inhabitants per hectare can provide powerful cues and insights.

4) Employing multiple archaeological proxies should be the rule of thumb when dealing with the reconstruction of past population dynamics. In this perspective, the convergence and divergence among the demographic trends inferred from various and independent demographic proxies can provide us with more powerful and robust insights. For example, the divergence among one or more trends can inform us about shifts in settlement patterns as evidenced above for central Italy during the transition between Final Bronze Age and Early Iron Age. In the case of settlement nucleation, people abandoned small and medium-sized settlements to concentrate on fewer and larger urban centres. Such a shift would have not been detected if in the analysis above I had not included the estimated size as a proxy (see Fig 8). Hence, the picture provided by the SPD of radiocarbon dates and the site count (and its derived proxies such as the aoristic sum and the randomised start date of sites) would be misleading and would indicate a decrease in population which, instead, boomed.

5) Open access practices. Scholars should make datasets and workflows (e.g., scripts written in R, Python, etc.) of their analyses freely accessible to guarantee the reproducibility of the research. This would make the academic community tighter and foster the exchange of knowledge and potential collaborations among scholars. Nowadays, scholars can benefit from a wide range of opportunities in the form of digital repositories on which is possible to deposit datasets free of charge (e.g., *Zenodo, figshare, Harvard Dataverse*), collaborative online

platforms (e.g., *Github*), and journals publishing data papers providing concise descriptions of datasets and their reuse potential (e.g., *Journal of Open Archaeology Data, Scientific Data*).

These recommendations should represent a basic starting point for those ones interested in palaeodemographic studies. Of course, some of them can be challenging to implement but they represent ideal guidelines to design data collection and research objectives. This chapter has focused only on the three most popular and widespread kinds of archaeological proxies but it does not exclude the use of other lines of evidence for inferring past population dynamics such as graves, count of houses, and surface artefact density. The choice of demographic proxies should be also examined carefully in the light of the research questions. For instance, a direct comparison between demographic proxies and palaeoenvironmental records (e.g., cave speleothems, lake sediments, pollen diagrams) may provide powerful cues about past human-environment interactions and more specifically about the human impact on the landscapes and human communities' resiliency under conditions of endogenous and exogenous stress. Even in this case, users should pay attention to the differing chronological resolutions and spatial distributions of the demographic and palaeoenvironmental records and choose the most suitable datasets for a direct comparison accordingly. Finally, the methodological advances made in the last decade have made a great contribution to palaeodemography, which has been recently benefitting from novel endeavours adopting Bayesian approaches to make robust inferences about past population trends[42].

[42] E. R. Crema, 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), 2021.