Radiocarbon, Vol 63, Nr 2, 2021, p 429–438 DOI:10.1017/RDC.2020.129

CLUSTERING OF CALIBRATED RADIOCARBON DATES: SITE-SPECIFIC CHRONOLOGICAL SEQUENCES IDENTIFIED BY DENSE RADIOCARBON SAMPLING

Peter Demján[1]* • Peter Pavúk[2]

[1]Institute of Archaeology of the Czech Academy of Sciences, Prague, Letenská 4, 11801 Praha 1, Czech Republic
[2]Institute of Classical Archaeology, Charles University, Prague, Nám. Jana Palacha 2, 11638 Praha 1, Czech Republic

ABSTRACT. Calibrated radiocarbon (^{14}C) determinations are commonly used in archaeology to assign calendar dates to a site's chronological phases identified based on additional evidence such as stratigraphy. In the absence of such evidence, we can perform dense ^{14}C sampling of the site to attempt to identify periods of heightened activity, separated by periods of inactivity, which correspond to archaeological phases and gaps between them. We propose a method to achieve this by hierarchical cluster analysis of the calibrated ^{14}C dates, followed by testing of the different clustering solutions for consistency based on silhouette coefficient and statistical significance using randomization. Separate events identified in such a way can then be regarded as evidence for distinct phases of activity and used to construct a site-specific sequence. This can be in turn used as a Bayesian prior to further narrow down the distributions of the calibrated ^{14}C dates. We assessed the validity of the method using simulated data as well as real-life archaeological data from the Bronze Age settlement of Troy. A Python implementation of the method is available online at https://github.com/demjanp/clustering_14C.

KEYWORDS: clustering, phasing, radiocarbon, randomization testing.

INTRODUCTION

Calibrated radiocarbon (^{14}C) determinations represent probability distributions of the calendar year in which the sampled organism died. This date is principally associated with a specific event, which can be inferred from additional evidence. In archaeology, such an event can be, e.g. a burial, dated by the time of death of the interred, or the occurrence of fire—intentional or accidental—dated by the time when seeds of cultural plants were carbonized. We can usually assign these events to spatial contexts, such as settlement pits or layers, from which we can infer their relative chronological relations, be it vertical or horizontal stratigraphy. Based on these, we can construct a site-specific sequence, or phasing, which is one of the desired results of an archaeological investigation of a settlement or burial site. In reality, due to deposition and site formation processes (e.g. Schiffer 1996), the majority of archaeological contexts are of secondary and even tertiary nature and do not always provide reliable additional evidence for the chronological relations of their contents. In the absence of such stratigraphic evidence, we cannot construct the temporal development/phasing of the site based on ^{14}C dates alone due to their probabilistic nature. Even if we do recognize gaps in a series of calibrated ^{14}C dates, we cannot readily interpret them as hiatuses or transitional periods in occupation. They could be caused by a sparsity of sampling, or be an effect of the calibration curve, as has been described by Rhode et al. (2014).

A good example can be a cultural layer containing evidence of human occupation in the form of fragmented pottery and carbonized plant remains, with no archaeological features, such as pits or house foundations, preserved. If we date the organic artifacts, and the resulting calibrated ^{14}C dates are wide apart from each other, we can interpret this as evidence for occupation of the site in two or more phases. Often, however, the probability distributions

*Corresponding author. Email: Peter.demjan@gmail.com

of the dates from such a context overlap and it is difficult to estimate, much less to statistically test, whether they represent events occurring continuously during a single period of occupation, or rather in multiple phases separated by periods of decreased activity or abandonment. In order to rigorously approach this problem, we need a method to calculate the closeness of two dates, i.e. the probability that they actually represent the same event. We further need to eliminate the possibility that the observed gaps between dates are a product of irregular sampling or artifacts caused by the calibration process.

We will examine the thesis that by performing cluster analysis of a set of ^{14}C dates, where their measure of dissimilarity is derived from the probability that they represent the same event, we can determine the optimal amount of separate events in time that would explain their distribution at a specified level of significance. Separate events identified in such a way can then be regarded as evidence for distinct phases of activity and used to construct a site-specific sequence. This can be in turn used as a Bayesian prior to further narrow down the distributions of the calibrated ^{14}C dates.

A mathematical solution to clustering of uncalibrated ^{14}C dates has been proposed by Wilson and Ward (1981). Their purely mathematical approach cannot be, however, applied to calibrated dates represented by non-normal probability distributions. Furthermore, their proposed method does not test whether the observed clustering could be caused by irregularities in sampling or fluctuations of the calibration curve. The non-normal distribution of calibrated ^{14}C dates also rules out any approach that would use a chi-squared or any other standard statistical test which assumes normally distributed values.

Methods for testing the statistical significance of hierarchical clustering of values were developed for use in the field of genetics (Liu et al. 2008; Huang et al. 2014; Kimes et al. 2017). Here, randomization is used to test whether a certain degree of clustering can be the product of chance. This method was also impossible to apply on calibrated ^{14}C dates, since probability distributions based on simulated ^{14}C ages where the calendar ages of the samples are randomly drawn from a normal distribution do not produce a dissimilarity matrix with normally distributed values, which is a requirement of this approach.

In the following text, we will present theoretical considerations regarding the definition of archaeological events and their temporal clustering. We will further propose a clustering method and a test of the statistical significance of the results. The validity of this approach will be tested using simulated data as well as real-life archaeological data. Finally, we will discuss possible limitations and applicability of the method in archaeological praxis. The proposed method has also been implemented as a Python 3.6 script and is available under the GNU General Public License (Demján 2020; Supplementary Material File 1).

MATERIALS AND METHODS

Theoretical Considerations

To meet the prerequisite of our thesis—perform a cluster analysis of ^{14}C dates—we need to calculate their dissimilarity matrix. As was already mentioned, the distance function of two calibrated dates should be based on the probability that they represent the same event.

We propose a definition of an archaeological event for the purposes of scientific dating as a period of human activity, during which dateable artifacts or ecofacts are deposited at such a frequency that it is impossible to recognize a time gap between them using any available dating method. Therefore, the summed probability distribution of dates originating from one event should be indistinguishable from a summed probability distribution of simulated dates with the same resulting mean and standard deviation generated randomly using the same dating uncertainty and calibration method. The reasoning behind this is that the dated artifacts or ecofacts represent a random sample of those continuously generated over the duration of the event. If the summed probability distribution of the observed dates significantly differs from the randomly generated versions, we must assume that we are dealing with more than one event. An underlying assumption for all considerations in this study is that the examined samples are spatially homogenous, i.e. the events that produced them affected the whole examined area. It is important to keep this in mind when interpreting the results.

Normality Test

As was established before, the non-normal nature of calibrated ^{14}C dates requires us to use randomization for statistical testing, which means comparing the observed dataset to a null model in which the actual calendar dates of the samples are normally distributed around a specific mean. To achieve this, the randomized summed probability distributions used for statistical testing must be generated using the same number of simulated dates as the observed dataset. The distribution of uncertainties of these dates must also be the same as the observed. Finally, the summed distribution must have the same mean and standard deviation as the observed summed distribution. This is to ensure that we take uncertainties due to measurement and calibration into account. The generator algorithm first produces an initial guess of randomized distributions by simulating ^{14}C dates of samples with calendar ages normally distributed with the mean and standard deviation of the observed summed distribution, randomly choosing uncertainty values from the observed dataset. This ^{14}C set is then optimized using the basinhopping function from the scipy.optimize package (Virtanen et al. 2020) to minimize the distance of the mean and standard deviation of the sum of the randomized distributions from the observed values. For a Python implementation of this algorithm, see function gen_random_dists (Demján 2020).

First, we need to test the null hypothesis (H_0) that the observed dates represent a single event, i.e. the optimal number of clusters explaining the evidence is one. We do this by generating a sufficient amount of sets of randomized distributions as described above and calculate the probability of achieving values as extreme as the observed ones for each calendar year. If at least one of these probabilities is lower than the significance level α, then H_0 has been rejected (Demján 2020, functions get_randomized, calc_percentiles) and we can assume that the dates form two or more clusters.

Clustering Method

The probability that two calibrated radiocarbon dates *i* and *j*, defined by mean radiocarbon ages t_i, t_j and standard deviations σ_i, σ_j, represent the same event can be expressed as the ratio

$$P_{ij} = \frac{4\sum_{t\in I} f_{Calib}(t, t_i, \sigma_i) f_{Calib}(t, t_j, \sigma_j)}{\left(\sum_{t\in I} f_{Calib}(t, t_i, \sigma_i) + \sum_{t\in I} f_{Calib}(t, t_j, \sigma_j)\right)^2} \quad (1)$$

where I is the set of all calendar dates from the IntCal13 (Reimer et al. 2013) and f_{Calib} is the calibration function defined by Bronk Ramsey (2008). It has to be noted that this calculation models the ideal case where the event's duration is one calendar year, so the resulting percentages will be very low for real-life data. This ideal model is used because we cannot

https://doi.org/10.1017/RDC.2020.129 Published online by Cambridge University Press

assume a specific time interval of the event and need this value only to calculate the chronological distance of two dates for the purposes of cluster analysis, without the need to directly interpret it. The distance function can then be defined as:

$$D_{ij} = 1 - P_{ij} \tag{2}$$

The next step towards cluster analysis is the calculation of a dissimilarity matrix of all calibrated dates using the function D_{ij} (2). For a Python implementation, see function calc_distance_matrix in the clustering_14C program (Demján 2020).

Further, we calculate Principal Component Analysis (PCA) scores for each row of the dissimilarity matrix and keep only components which explain 99% of the variance (Demján 2020, function calc_distances_pca). This step is especially important for larger datasets to reduce the number of dimensions (i.e. "de-noise" the data) and achieve a more stable clustering. Alternative methods of choosing the number of components to keep were suggested by one of the reviewers of this paper, one being to discard just the last component, which can be assumed to be the one capturing noise, the other being Horn's Parallel Analysis, which selects components based on randomization testing (Horn 1965). We tested both approaches on simulated data and while Horn's method always rejected all but the first component, which led to a substantial reduction of fidelity, the selection of all but the last component produced comparable results with the 99% variance method (see Supplementary Material File 2).

Hierarchical Cluster Analysis (HCA) is then performed on the PCA scores using Ward's method based on the Euclidean distance metric. Solutions are produced for numbers of clusters ranging from two to number of dates minus two (Demján 2020, function calc_clusters_hca).

Statistical Testing of Clustering Solutions

If the normality test rejects H_0 for a one-cluster solution, we can proceed with testing the HCA solutions with two and more clusters. For each solution, we calculate the mean Silhouette Coefficient, which quantifies the consistency of clustering (Rousseeuw 1987). We then test the H_0, that such consistency can be achieved by clustering datasets generated randomly under the previously described conditions (Demján 2020, function get_clusters_hca). Out of the solutions for which H_0 was rejected, the optimal is then the one with the highest mean silhouette coefficient, i.e. the most consistent one (Demján 2020, function get_opt_clusters).

RESULTS

The viability of the clustering method proposed in this paper depends on its ability to detect separate events in archaeological data. There are two possible strategies to test this ability. The first is using simulated data where we know the exact amount of events to be detected if the method is successful. The second is using actual archaeological data where a sufficient amount of additional evidence allows us to assign them to specific events even without ^{14}C dating.

Simulated Data

Using the clustering method on simulated data allows us to rigorously test its viability and also estimate conditions under which it is applicable, given different levels of uncertainty due to the shape of the calibration curve in different time periods. We simulated ^{14}C dates of events occurring in calendar dates from 6000 years BC to 1500 years AD, with dating uncertainties randomly picked from an interval of 10–30 ^{14}C years. For each simulated

https://doi.org/10.1017/RDC.2020.129 Published online by Cambridge University Press

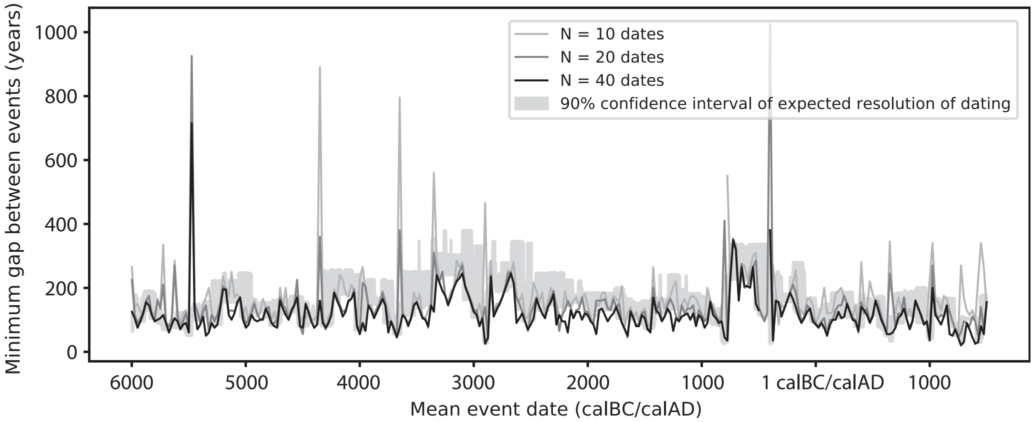

Figure 1 Minimum gap between simulated events necessary to correctly detect them using clustering of calibrated ^{14}C dates. Expected resolution of dating according to Svetlik et al. (2019) shown for comparison.

observation, we generated versions with 10, 20, and 40 dates, divided between three events. The simulated calendar ages of the events had a mean around a specific year in the 6000 BC to 1500 AD range and were separated by a certain time gap. For each simulated dataset, we calculated the optimal number of clusters. The time gap was gradually increased until the calculated number of clusters equalled the simulated number of three events. The detection of the correct number of clusters for a specific mean calendar age of the events and time gap had to be successful over 10 successively generated datasets to be considered stable. Figure 1 shows the minimum time gap in calendar years between archaeological events which occurred around a certain calendar date to successfully detect them based on a specific number of ^{14}C determinations.

The results show that events separated by under 100 years can be detected in optimal parts of the calibration curve and even in the less optimal parts, the necessary gap does not usually exceed 300 years. In general, it can be said that the expected resolution of clustering of calibrated ^{14}C dates is similar to the expected resolution of the dating itself, as reported by Svetlik et al. (2019).

Archaeological Data: Troy

To test the new method on real-life data, we selected the Bronze Age settlement of Troy (Western Anatolia) as an ideal candidate. It is a well-known site with complex stratigraphy (ca. 3000 BC–500 AD), that has been excavated over many years, with sufficient publications (for summaries cf. Blegen 1963; Korfmann 2006; Pernicka et al. 2014) and a robust set of published ^{14}C dates (Korfmann and Krommer 1993; Krommer et al. 2003). As the understanding of the stratigraphy and the various depositional processes at Troy has moved on since the initial publication of the ^{14}C data, the present paper targets only the 40 ^{14}C dated samples from the Middle and Late Bronze Age at Troy (roughly 2nd millennium BC), whose stratigraphy and relative dating were more recently reassessed (Pavúk 2014; Pavúk 2020). These include mainly charcoals but also seeds and human collagen from well-stratified contexts, reflecting the range and frequency of the available samples. Most were taken in the 1990s and early 2000s and the sampling strategy was not as systematic as one would hope for today. The chronological development of this site in the second millennium BC is, nevertheless, well documented and the depositional processes understood by now,

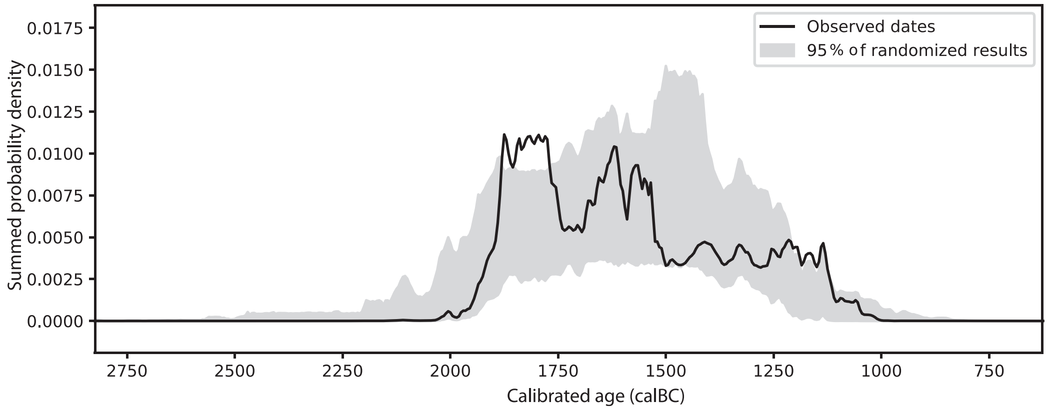

Figure 2 Summed probability distributions of calibrated ^{14}C dates from Troy compared to an acceptance envelope of summed probability distributions of randomized dates with a significance level α=0.05.

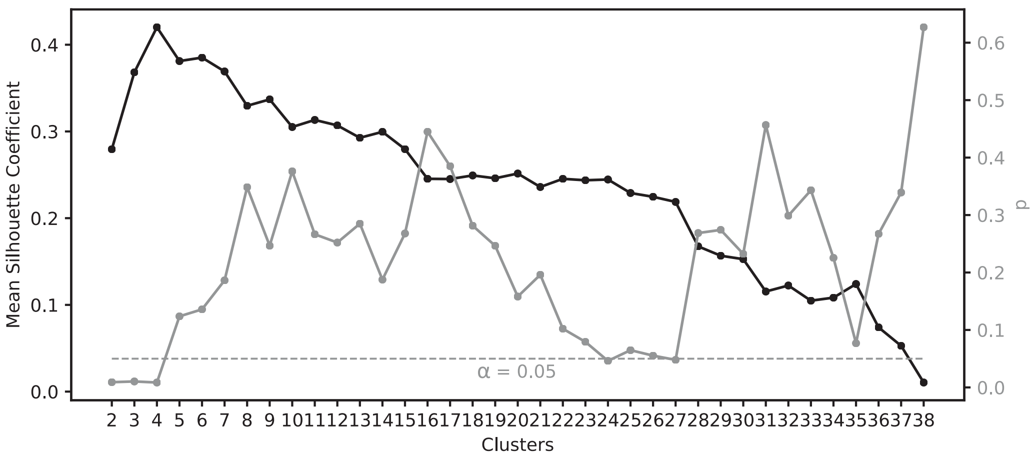

Figure 3 Mean silhouette coefficient (black line) of different clustering solutions of calibrated ^{14}C dates from Troy with corresponding p-values (grey line). Significance level α=0.05 is denoted by a dashed line.

including any caesurae in occupation due to catastrophic events or rebuilding. We can thus compare the chronological phasing derived from clustering calibrated ^{14}C dates with the already archaeologically established phasing for the site.

The H_0 that the dates represent a normal distribution was rejected at the significance level α=0.05 (Figure 2). The optimal clustering solution was subsequently found to be 4 clusters with ca 100-year gaps between them (Figure 3; Supplementary Material Files 3–6).

The first thing we noticed was that while the clusters do not correlate with the individual architectural phases identified so far, they almost perfectly match the overall periodization established for Troy: cluster 1 is the Middle Bronze Age (MBA), cluster 2 is Late Bronze (LB) 1, and clusters 3 and 4 correspond to LB 2A and LB 2B at Troy (Figure 4, for the

https://doi.org/10.1017/RDC.2020.129 Published online by Cambridge University Press

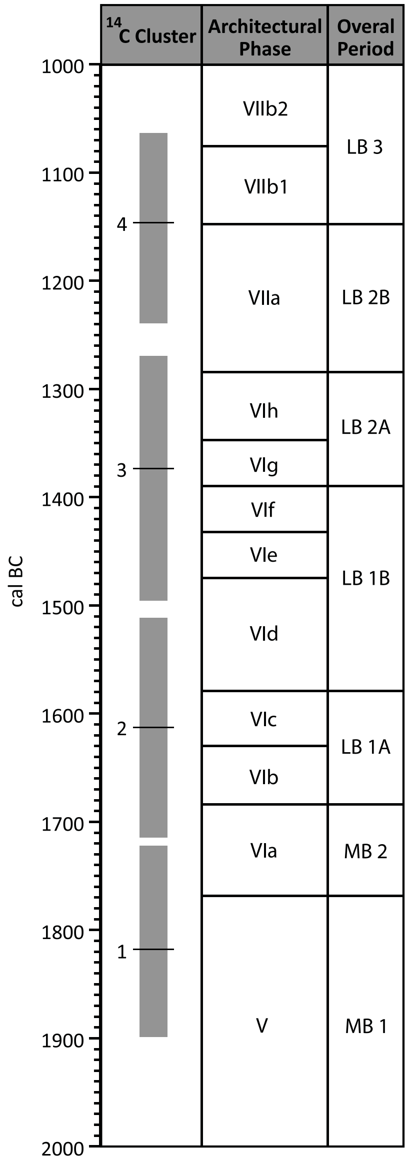

Figure 4 ^{14}C clusters matched against periodization of the Middle (MB) and Late Bronze Age (LB) Troy. The box plots represent mean and maximum extent of 95.4% calibrated age ranges of each cluster.

periodization, see Pavúk 2020, cf. also Pavúk and Horejs 2018). What was also distinctive were the gaps between the clusters, to whose interpretation we shall turn now.

What comes into mind in the case of Troy are the major disruptions or reshapings of the citadel and its vicinity. Here, we speak not about single-deposit formation processes, but large-scale site-formation processes. The gaps between the clusters correspond exactly to these: Whereas

https://doi.org/10.1017/RDC.2020.129 Published online by Cambridge University Press

Troy V and Earliest Troy VI (cluster 1 with a midpoint around 1830 BC) have similar formation processes and gradual accumulation, it is in the latter part of Troy VI-Early and in VI-Middle (cluster 2 with midpoint around 1630 BC) that we see the first attempts at the intentional reshaping of the citadel, with the construction of new terraces and terrace walls, for instance. This reshaping is developed more extensively during Troy VI-Late (cluster 3 with midpoint around 1400 BC), which ends in a major earthquake. This is followed by an almost complete rebuilding of the citadel and also adjacent areas during Troy VIIa, followed by similar activities also in Troy VIIb1 (both cluster 4 with midpoint around 1190 BC). Thus, what we see are disruptions in the otherwise gradual accumulation of deposits, which would rather support a normative distribution of burning of habitational events (as likely happened in real life at Troy).

To test the robustness of the solution, we calculated optimal clustering for 10 randomly selected subsets, half the size of the original dataset (20 dates). Out of these, 7 solutions had a number of clusters less than half the number of samples (i.e. a cluster contained more than two samples on average). Cluster boundaries in these solutions were at the same places as in the full set, meaning that even clustering a randomly selected half-sized dataset correctly detected the chronological phases in 7 out of 10 cases (Supplementary Material File 3).

DISCUSSION AND CONCLUSIONS

We have successfully demonstrated on simulated data that it is possible to determine the number of events from which ^{14}C dates originate by the proposed method of hierarchical cluster analysis and randomization testing, given that they are separated by sufficiently long gaps and the sampling is dense enough. This approach was further validated by correctly detecting major phases in the development of the Bronze Age settlement of Troy, which are already known based on previous archaeological and architectural research.

Possible issues preventing a successful application of the new method can arise due to sampling density or limitations of ^{14}C dating itself. If the number of samples is too low with respect to the length of the examined time period, the optimal solution can result in too many clusters, each consisting of one or two dates. In such a case, we would be probably detecting gaps in collected evidence, rather than actual gaps in past activities that produced it (see clustering of subsets 2, 6 and 7 in Supplementary Material File 3). Another issue could arise if the gaps between events are shorter than the temporal resolution of ^{14}C dating, be it due to uncertainty of the measurement, or a plateau in the calibration curve resulting in a wide spread of the probability distributions of calendar ages (see Figure 1).

The method is especially suitable for cases in which there is little or no archaeological evidence for a chronological development of a site, i.e. we are unable to derive chronology from stratigraphic relations of archaeological features or typological evaluation of the artifacts but still have sufficient evidence of human activity in form of ^{14}C dateable macro-remains (e.g. charred cereals, wood from fireplaces, etc.). An example of such a study with the application of a preliminary version of the method presented here is a paper by Dreslerová et al. (2020). Bayesian modelling can then be used to further infer on the duration of gaps between the detected clusters, such as the Interval function in OxCal. Applying clustering as a prior can also to a lesser degree affect the modelled span of a sequence of dates. For an example, comparing OxCal models of the same set of 20 dates originating from two simulated events 100 years apart with and without clustering, see Supplementary Material Files 7–9.

Another use could be with data from stratigraphically disconnected but still spatially close areas of research (e.g. different excavation trenches at the same archaeological site), where we could detect periods of activity (e.g. phases of occupation) based on ^{14}C dates, even if we are unable to chronologically link those areas based on other evidence.

Clustering could also to some extend provide a solution to temporal "binning" of ^{14}C dates from large datasets where it is either impossible or impractical to combine them based on archaeological evidence. The requirement to ensure spatial homogeneity of the binned data (i.e. that they can actually represent one phase of a site) of course still applies. Especially for inference on spatio-temporal settlement density, an interpolation method incorporating also the spatial dimension, such as Evidence Density Estimation (Demján and Dreslerová 2016; Demján 2019) might be more appropriate.

Even if a relative chronological sequence is already known, as is the case of the example from Troy presented here, we can use the randomization and clustering approach to determine whether the occupation of the site was continuous, or there are detectable gaps. In the examined case, a possible interpretation of those gaps would be an overall change in site formation processes, be it destruction or major rebuilding programs. For further stimulating thought concerning gaps and Troy see Weninger and Easton (2014), who target the EBA sequence.

The possibility to infer chronological sequences based purely on ^{14}C dates demonstrates that under certain conditions, the mere fact that samples are of archaeological origin (i.e. they are a product of human activities) and are spatially homogenous can be used as a Bayesian prior to increase the precision of their absolute dating.

ACKNOWLEDGMENTS

This publication was supported by OP RDE, MEYS, under the European Regional Development Fund-Project "Ultra-trace isotope research in social and environmental studies using accelerator mass spectrometry," Reg. No. CZ.02.1.01/0.0/0.0/16_019/0000728 in collaboration with the European Regional Development Fund-Project "Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World," Reg. No. CZ.02.1.01/0.0/0.0/16_019/0000734. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/RDC.2020.129

REFERENCES

Blegen CW. 1963. Troy and the Trojans. Ancient peoples and places 32. New York: Thames and Hudson.

Bronk Ramsey C. 2008. Radiocarbon dating: revolutions in understanding. Archaeometry 50(20):249–275. doi: 10.1111/j.1475-4754.2008.00394.x.

Demján P. 2019. EDE interpolation. QGIS Python plugins repository [accessed 2020 Aug 07]. https://plugins.qgis.org/plugins/ede_interpolation.

Demján P. 2020. Clustering_14C - Clustering of calibrated radiocarbon dates [accessed 2020 Mar 03]. https://github.com/demjanp/clustering_14C.

Demján P, Dreslerová D. 2016. Modelling distribution of archaeological settlement evidence based on heterogeneous spatial and temporal data. Journal of Archaeological Science 69:100–109. https://doi.org/10.1016/j.jas.2016.04.003.

Dreslerová D, Kozáková R, Metlička M, Brychová V, Bobek P, Čišecký Č, Demján P, Lisá L, Pokorná A, Michálek J, et al. 2020. Seeking the meaning of a unique mountain site through a multidisciplinary approach. The Late La Tène site at Sklářské Valley, Šumava Mountains, Czech Republic. Quaternary International, 2020 preprint. doi: 10.1016/j.quaint.2020.03.013.

Horn J. 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 32:179–185.

Huang H, Liu Y, Yuan M, Marron JS. 2014. Statistical significance of clustering using soft thresholding. Journal of Computational and Graphical Statistics 24(4):975–993. doi: 10.1080/10618600.2014.948179.

Kimes PK, Liu Y, Hayes DN, Marron JS. 2017. Statistical significance for hierarchical clustering. Biometrics 73(3):811–821. doi: 10.1111/biom.12647.

Korfmann MO, editor. 2006. Troia. Archäologie eines Siedlungshügels und seiner Landschaft. Mainz: Zabern.

Korfmann M, Kromer B. 1993. Demircihüyük, Beşik-Tepe, Troia. Eine Zwischenbilanz zur Chronologie dreier Orte in Westanatolien. Studia Troica 3:135–171.

Kromer B, Korfmann M, Jablonka P. 2003. Heidelberg Radiocarbon Dates for Troia I to VIII and Kumtepe. In: Wagner GA, Pernicka E, Uerpmann HP, editors. Troia and the Troad. Scientific Approaches. Berlin: Springer. p. 43–54. doi: 10.1007/978-3-662-05308-9_4.

Liu Y, Hayes DN, Nobel A, Marron JS. 2008. Statistical significance of clustering for high-dimension, low-sample size data. Journal of the American Statistical Association 103(483): 1281-1293. doi: 10.1198/016214508000000454.

Pavúk P. 2014. Troia VI Früh und Mitte. Keramik, stratigraphie, chronologie. Studia Troica Monographien 3. Bonn: Habelt.

Pavúk P. 2020. Relative and absolute chronology: Troy VI and VII. In: Pieniążek M, Pavúk P, Thumm-Doğrayan D, Pernicka E, editors. Troia 1987–2012: Grabungen und Forschungen III. Die Spätbronzezeitlichen Schichten. Studia Troica Monographien 7. Bonn: Habelt. Forthcoming.

Pavúk P, Horejs B. 2018. Ceramics, surveys, and connectivity in western Anatolia: the Middle and Late Bronze Age Bakırçay/Kaikos Valley restudied. Ägypten und Levante 28:457–485. doi: 10.1553/AEundL28s457.

Pernicka E, Rose CB, Jablonka P, eds. 2014. Troia 1987–2012: Grabungen und Forschungen I. Forschungsgeschichte, Methoden und Landschaft. Studia Troica Monographien 5. Bonn: Habelt.

Reimer PJ, Bard E, Bayliss A, Beck JW, Blackwell PG, Ramsey CB, Buck CE, Cheng H, Edwards RL, Friedrich M, et al. 2013. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. Radiocarbon 55(4): 1869–1887. doi: 10.2458/azu_js_rc.55.16947.

Rhode D, Brantingham PJ, Perreault C, Madsen DB. 2014. Mind the gaps: testing for hiatuses in regional radiocarbon date sequences. Journal of Archaeological Science 52:567–577. doi: 10.1016/j.jas.2014.02.022.

Rousseeuw PJ. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20(11):53–65. doi: 10.1016/0377-0427(87)90125-7.

Schiffer MB. 1996. Formation processes of the archaeological record. Salt Lake City: University of Utah Press.

Svetlik I, Jull AJT, Molnár M, Povinec PP, Kolář T, Demján P, Pachnerova Brabcova K, Brychova V, Dreslerová D, Rybníček M, et al. 2019. The best possible time resolution: how precise could a radiocarbon dating method be? Radiocarbon 61(6):1729–1740. doi: 10.1017/RDC.2019.134.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods 17(3):261–272. doi: 10.1038/s41592-019-0686-2.

Weninger B, Easton D. 2014. The Early Bronze Age Chronology of Troy (Periods I-III): Pottery Seriation, Radiocarbon Dating and the Gap. In: Horejs B, Mehofer M, editors. Western Anatolia before Troy: proto-urbanisation in the 4th millennium BC? Proceedings of the International Symposium held at the Kunsthistorisches Museum Wien, Vienna, Austria, 21–24 November, 2012. Oriental and European Archaeology 1, Vienna: Austrian Academy of Sciences Press. p. 157–199.

Wilson SR, Ward GK. 1981. Evaluation and clustering of radiocarbon age determinations: procedures and paradigms. Archaeometry 23(1):19–39. doi: 10.1111/j.1475-4754.1981.tb00952.x.