Journal of Archaeological Science 87 (2017) 1–9

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: http://www.elsevier.com/locate/jas

Spatio-temporal approaches to archaeological radiocarbon dates

E.R. Crema [a, *], A. Bevan [b], S. Shennan [b]

[a] Department of Archaeology and Anthropology, University of Cambridge, Cambridge, CB2 3ER, UK
[b] UCL Institute of Archaeology, London, WC1H 0PY, UK

ARTICLE INFO

Article history:
Received 17 July 2017
Received in revised form
30 August 2017
Accepted 4 September 2017

Keywords:
Summed probability distribution of radiocarbon dates
Prehistoric demography
Spatial analysis
European Neolithic
Simulation

ABSTRACT

Summed probability distributions of radiocarbon dates are an increasingly popular means by which to reconstruct prehistoric population dynamics, enabling more thorough cross-regional comparison and more robust hypothesis testing, for example with regard to the impact of climate change on past human demography. Here we review another use of such summed distributions – to make spatially explicit inferences about geographic variation in prehistoric populations. We argue that most of the methods proposed so far have been strongly biased by spatially varying sampling intensity, and we therefore propose a spatial permutation test that is robust to such forms of bias and able to detect both positive and negative local deviations from pan-regional rates of change in radiocarbon date density. We test our method both on some simple, simulated population trajectories and also on a large real-world dataset, and show that we can draw useful conclusions about spatio-temporal variation in population across Neolithic Europe.

1. Introduction

The last decade has seen a rapid increase in the collection and analysis of radiocarbon dates to infer long-term changes in human population density. Early efforts based on the visual inspection of time-series generated from small numbers of uncalibrated dates (Rick, 1987; Ames, 1991) are now being replaced by statistical analysis of thousands of dates, shedding new light on long-term prehistoric population change. The most widely adopted approach is the *summed probability distribution of radiocarbon dates* (SPDRD), and its success is rooted both in the increasing availability of large collections of ^{14}C dates (e.g. Gajewski et al., 2011; Williams et al., 2014; Manning et al., 2016; Chaput and Gajewski, 2016 for a review) and in its enabling of detailed time-series comparisons that are not possible via other demographic proxies such as settlement counts. By building a time-series that is based on an absolute, high resolution chronology (but see Weninger et al., 2015 for the perils of placing too much faith in chronological resolutions below 200 yrs), the SPDRD makes it possible to directly compare inferred patterns of prehistoric population change to paleo-climatic reconstructions (e.g. Shennan et al., 2013; Kelly et al., 2013), and to attempt large scale, cross-cultural and cross-regional analysis (Chaput and Gajewski, 2016; Crema et al., 2016; Zahid et al., 2016).

SPDRD-based methods, however, have been subject to criticism with respect to a potentially wide variety of biases that might produce spurious patterns in the radiocarbon time series, and therefore lead to misleading conclusions about population change. These biases include: 1) sampling error (Timpson et al., 2014); 2) idiosyncrasies associated with the calibration process (Weninger et al., 2015); 3) time-dependent taphonomic loss (Surovell and Brantingham, 2007); 4) spatial and/or temporal differences in site-to-population ratios; and 5) spatial and/or temporal differences in sampling intensity. Nevertheless, we would argue that most of these problems have been partially or fully overcome. The effects of calibration wiggles and sampling error have been approached by examining SPDRDs in relation to one or more null models (Shennan et al., 2013; Timpson et al., 2014; Porčić, and Nikolić, 2016; Crema et al., 2016; Goldberg et al., 2016), rather than simply qualitatively assessing time-series. The problem of taphonomic loss has been addressed through the development of correction formulae (Surovell et al., 2009; Kelly et al., 2013), even if many researchers still prefer working with data that have not been adjusted in this way. Differences in site-to-population ratios have been modelled using ethnographic data (Downey et al., 2014), while spatial differences in research intensity have been tackled by means of non-parametric permutation tests (Crema et al., 2016). While these solutions are certainly not universal and do not overcome other potential sources of bias (e.g. within-region temporal

* Corresponding author.
E-mail address: erc62@cam.ac.uk (E.R. Crema).

http://dx.doi.org/10.1016/j.jas.2017.09.007
0305-4403/

differences in sampling intensity, spatially uneven taphonomic and deposition processes), they do show how ad hoc analytical solutions can improve the inferential power offered by an SPDRD.

Given the popularity of these techniques, it is not surprising that some researchers wish to move beyond a single global time-series analysis and consider local, spatially sensitive versions as well. In broad terms, attempts at such spatio-temporal analysis of radiocarbon dates can be grouped into two categories. The first category includes several studies that compare the SPDRD of multiple regions (e.g. Wang et al., 2014, Bernabeu Aubán et al., 2016, Crema et al., 2016; Miller and Gingerich, 2013, Shennan et al., 2013; Timpson et al., 2014; Gayo et al., 2015) either visually or by means of some statistical testing (as in Crema et al., 2016). The second approach is more explicitly spatial, as it is not based on *a priori* subdivision of space into subregions, and instead seeks to reconstruct continuous changes (in space and time) in the density of radiocarbon dates (Collard et al., 2010; Grove, 2011; Chaput et al., 2015; Goldberg et al., 2016; Manning and Timpson, 2014; Perez et al., 2016; Onkamo et al., 2012). In most cases this involves using some form of kernel density estimate (KDE) with the summed probability at a given location (site) and time-slice being treated as a weight (e.g. Collard et al., 2010; Grove, 2011; Manning and Timpson, 2014; García Puchol et al., 2017). Perez et al. (2016) offer an alternative solution based on inverse distance weighting and per raster-cell site frequency. They chose a coarse chronological resolution of 2000 years, which effectively treats each site as a binary presence/absence, but their approach can, at least in principle, be applied to finer temporal intervals, substituting site frequencies for summed probabilities. Onkamo et al. (2012) adopt a more sophisticated solution using a hierarchical Bayesian model with a conditional autoregressive model. Their solution is robust to the problem of sparse data (see below) and hence ideal in many archaeological contexts, but their implementation requires the use of binary data (i.e. presence/absence), which they obtained by examining whether the posterior mean of the calibrated estimate of each ^{14}C date was within predefined discrete chronological phases. The solution adopted by Onkamo and colleagues is broadly in line with similar approaches to spatial analysis that quantify the effect of environmental covariates (e.g. Bevan et al., 2013; Eve and Crema, 2014), and useful in cases of coarser chronological resolution where the way each calibrated date is assigned to one phase or another only has a minor influence on the analytical outcome.

This incipient interest in spatio-temporal analysis of ^{14}C dates does however raise a series of new challenges. First, most of the methods proposed so far (except for Onkamo et al., 2012) generate density maps that are primarily visualisations, usually without any accompanying formal assessment of whether a specific pattern is genuine or a spurious artefact of sampling error. This is exactly the same problem that led to the development of hypothesis-testing approaches in the analysis of SPDRD in the first place (see Timpson et al., 2014), except that the transition from discrete regions to a continuous space entails a substantial reduction in the per unit sample size. Thus, other things being equal, spatio-temporal analysis of ^{14}C dates should be affected more strongly by sampling error bias, and we should therefore be *more* rather than less cautious about what we are visualising. Second, increasing the geographic extent of the analysis will also increase the chance of a spatially heterogeneous sampling intensity of radiocarbon dates. A quick glance at many of the large datasets of radiocarbon dates shows spatial differences in the density of ^{14}C dates that are remarkably in line with modern state boundaries (e.g. Wyoming in the CARD dataset or Ireland in the EUROEVOL dataset), reflecting regional diversity in research traditions. This is certainly not a new problem in archaeology (see Hodder and Orton, 1976; Fitzpatrick, 1987 for early discussions), and is common to other fields of study (e.g. Syfert et al., 2013, Stolar and Nielsen, 2015 for problems and solutions in ecological models of species distributions). Chaput et al. (2015) have recently proposed a solution applicable for spatial density estimates of radiocarbon dates. Their solution consists of generating a sampling intensity map using all the site locations (thus irrespective of their temporal stamp) and then dividing the weighted KDE of each temporal slice by this map, effectively de-trending the spatial variation in research intensity. While the density maps are still assessed exclusively in visual fashion, the approach proposed by Chaput and colleagues solves an often-neglected problem in the spatio-temporal analysis of archaeological data in general.

In this paper, we contribute to this growing research agenda by introducing a new technique based on local spatial statistics (Getis and Ord, 1996; Premo, 2004; Crema and Bianchi, 2013 for some archaeological examples), that combines the hypothesis-testing approach proposed by Shennan et al., (2013) and the permutation-based null model detailed in Crema et al. (2016). This new method can determine whether there are any locally and statistically significant positive or negative deviations — which we will respectively refer to as *hot* and *cold spots* — from the null model. The null model in this case is represented by the global rate of growth within the study region as a whole. Thus, if there are no spatial differences in the rate of increase or decrease in the density of ^{14}C dates we should not expect to observe any *hot* or *cold spots*. The presence of *hot* or *cold spots* would conversely indicate that some regions experienced a higher (*hot spots*) or lower (*cold spots*) rate of growth compared to the global trend.

With the above goals in mind, this paper is structured as follows. Section 2.1 below first details the proposed method, for which source code and sample scripts can be found in the supplementary information. We then test our method on: 1) a simulated dataset where differences in sampling intensity and spatio-temporal patterns of population density are known, and 2) radiocarbon dates associated with Neolithic Europe, using the EUROEVOL database (Manning et al., 2016). We then discuss the results, considering the benefits and the limits of the proposed method.

2. Materials and methods

2.1. Methods

The core steps of the method proposed here consist of: 1) developing a proxy for local demography by means of a weighted SPDRD; 2) computing rates of growth within defined chronological intervals; and 3) comparing local rates to a global one (the null model) thereby identifying episodes of positive (*hot spots*) and/or negative (*cold spots*) deviations. The program code, written in R statistical computing language (R Core Team, 2016), is available online as an R package (https://github.com/ahb108/rcarbon), while the electronic supplementary material contains the scripts to reproduce the results in the manuscript and a brief tutorial (see also https://github.com/ercrema/JAS2017_spatialSPD).

2.1.1. Generating a locally weighted SPD

Given a data set consisting of $i=1,2,\ldots L$ site locations, a "local" SPDRD for each site location i is computed in two sequences of steps. Firstly, a *site-level* SPDRD is computed as follows:

1. Samples associated with each site i are "binned" (i.e. grouped) based on the Euclidean distance of their radiocarbon age using a complete linkage agglomerative hierarchical method with a cut off value of k years.

2. Samples associated with each bin are then calibrated, and their probabilities summed and normalised to unity.
3. The results are then summed at the site level, so that the sum of the probabilities associated with each site is equal to its number of bins.

These steps are equivalent to the SPDRD proposed in previous studies (Shennan et al., 2013; Timpson et al., 2014). The locally weighted SPDRD consists of combining the probability of neighbouring sites as follows:

4. For each possible pair of locations i and j calculate the weight $w_{i,j}$ with the following equation (Fotheringham et al., 2000:111):

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{h^2}\right) \quad (1)$$

where d_{ij} is the great arc distance between i and j and h is a user-defined Gaussian kernel bandwidth. This is a widely used formula (though alternative distance decay functions are available, see Fotheringham et al., 2000) where the contribution of neighbouring sites declines fractionally and gradually as a function of distance.

5 Define T temporal slices and sum the probabilities within each interval for each location i
6. For each location i and each temporal slice compute the weighted sum of all probabilities resulting from step 5 using the weight obtained in step 4:

$$wSPDRD_{i,t} = \sum_j^L w_{ij} \cdot SPDRD_{j,t} \quad (2)$$

where $SPDRD_{i,t}$ is the summed probability of radiocarbon dates at location i at time-block t. It is worth reminding that the focal site i is included in the set L of all sites, and given eq.(1) its contribution the summed probability is not fractional (as w_{ij} reduces to unity when $d_{ij} = 0$).

The result of steps 1–6 will generate L local weighted SPDRDs, with the probabilities assigned to neighbouring sites added to the one associated to a given focal site i. The exact contribution of neighbouring sites depends on their distant to i; sites in proximity will have greater weights, whilst sites located at larger distances from i will have virtually no contribution to the local SPD of i.

2.1.2. Estimating rates of growth between temporal slices

The next step is the calculation of the geometric growth rate at each location for each abutting pair of temporal slices. This is given by the following equation (cf. Brown, 2017, eq. (13)):

$$r_t = \left(\frac{wSPDRD_{t+1}}{wSPDRD_t}\right)^{\frac{1}{\Delta t}} - 1 \quad (3)$$

where $wSPDRD_t$ and $wSPDRD_{t+1}$ are the local weighted SPDRDs at two abutting time slices and Δt is the length (duration in time) of each slice. The final result will thus be a vector with $T-1$ rates of changes for each of the L site locations.

2.1.3. Permutation test

The permutation algorithm consists of randomly shuffling the locations (i.e. sites) associated with each local SPDRD, before executing the application of the spatial weights and computing the local growth rate. This process is iterated n times, so that for each location i, at each transition t to $t + 1$, there is an observed local growth rate $O_{i,t}$ and a vector $S_{1,i,t}$, $S_{2,i,t}$ … $S_{n,i,t}$ of simulated growth rates generated from the random permutation.

2.1.4. Computing significance of hot/cold spots

We define as *hot spots* all locations exhibiting an observed local growth rate that is significantly higher than the distribution of simulated local growth rates. It follows that *cold* spots are locations where the observed local growth rate is lower than the randomised set. P-values for both are computed using the formula $(r + 1)/(n+1)$ (North et al., 2002), where n is the total number of simulations, and r is the number of replicates where the $O_{i,t} \geq S_{i,t}$ (for the p-value p_{high} in the *hot spots*) or $O_{i,t} \leq S_{i,t}$ (for the p-value p_{low} in the *cold spots*).

Given that in most cases there are large numbers of site locations, there is a high risk of type I error (incorrectly rejecting a true null hypothesis) due to multiple-testing. As for several geostatistical analyses (cf. Anselin, 1995), p-value adjustment methods such as Bonferroni's correction are in this case too conservative and inflate type II errors (failing to reject a false null hypothesis). Here we approach the problem in terms of false discovery rate (Benjamini and Hochberg, 1995, 1997) computing the q-values q_{low} and q_{hi} for each location. While a p-value of 0.05 implies that 5% of the tests will result in false positives, a q-value of 0.05 means that 5% of the results that have a q-value less than 0.05 are false positive.

2.2. Materials

We examined our new method by using (1) an artificially generated data-set with a known pattern of spatial heterogeneity and known population trajectories through time and (2) an archaeological case study from Neolithic Europe, using the EUROEVOL dataset (Manning et al., 2016).

2.2.1. Simulated data

We considered a 40 × 40 square shaped study area divided into four equally sized sub-regions *A, B, C,* and *D* and a temporal interval of 7000 to 3000 BP. We then assumed three distinct population dynamics (see Fig. 1-a and 1-b): a rise and fall pattern (for region *B*), a delayed population increase (for region C); and an intermediate and steady population growth (for regions *A* and *D*). Based on these trajectories we then generated 5000 data points across space and time, aggregating coordinates to the nearest integer. We then converted the calendar dates into radiocarbon dates through back-calibration using an error estimate randomly drawn from a uniform distribution between 20 and 60 years. The resulting data set consisted of 5000 radiocarbon dates at 1387 unique locations in space. We also created a second set where half of the samples in regions B and D were randomly removed, mimicking thus a spatially heterogeneous sampling intensity.

The spatial permutation tests were executed for both simulated datasets, using: a 1) temporal bin size k of 200 (i.e. $k = 200$); a 2) spatial bandwidth of 6 units (i.e. $h = 6$); 3) temporal slices of 500 years (i.e. $\Delta t = 500$); and 4) 10,000 permutations to calculate the local p and q-values. Dates have been calibrated and back-calibrated using IntCal13 (Reimer et al., 2013), and were not normalised for the creation of the SPDRDs, following Weninger et al., 2015 (although the high setting of Δt reduces the difference between normalised and unnormalised SPDRDs).

2.2.2. Neolithic Europe

Radiocarbon dates from Europe have been extensively used in the past to infer population change (Gamble et al., 2005; Shennan and Edinborough, 2007; Collard et al., 2010; Shennan et al., 2013; Timpson et al., 2014), often sparking debates on whether they constitute a valid proxy for population change in the Neolithic or not (see e.g. Contreras and Meadows, 2014, Mökkönen, 2014, Tallavaara et al., 2014, Torfing, 2015, Timpson et al., 2015 for

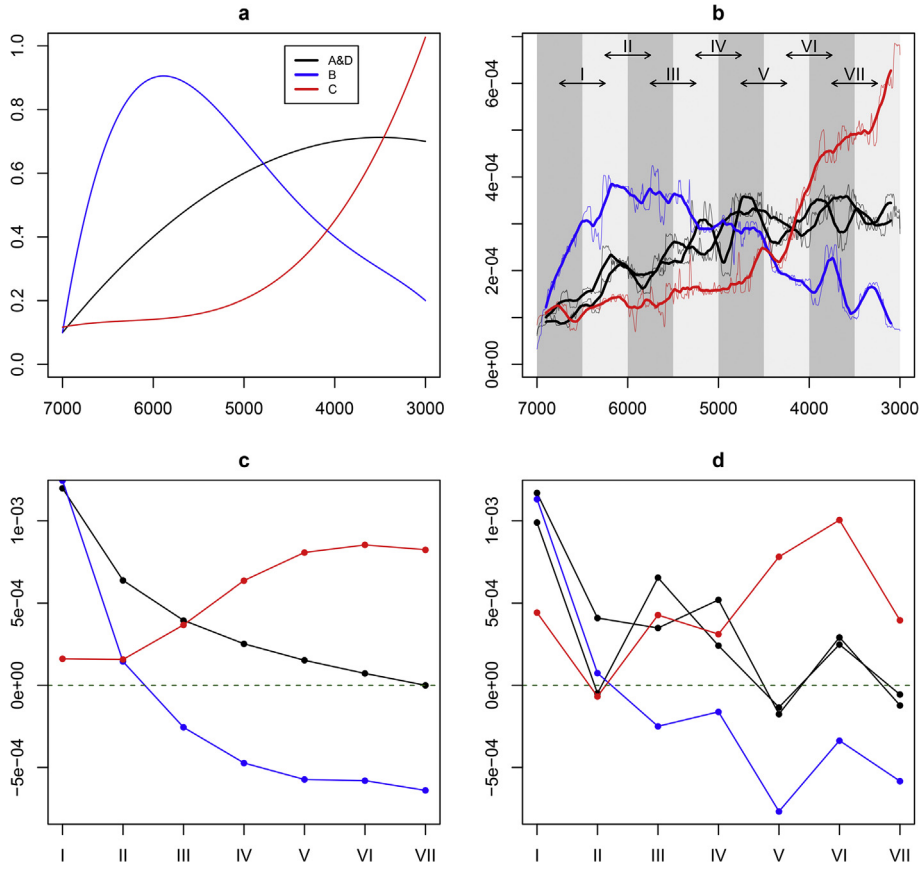

Fig. 1. *a*) relative change of population density across time in the four regions; *b*) SPDRD obtained from sample ^{14}C dates randomly drawn with probabilities proportional to the population density depicted in panel *a*; c) expected rate of growth for each transition (cf panel *b*) based on the population trajectories shown in panel *a*; d) observed rate of growth computed from the SPDRD (cf panel *b*).

arguments on both sides, see also similar debates for prehistoric Australia by Attenbrow and Hiscock, 2015, Hiscock and Attenbrow, 2016, Williams and Ulm, 2016, and Smith, 2016). Here we use the recently published EUROEVOL dataset (Manning et al., 2016), which has been extensively examined at various regional scales (see Shennan et al., 2013; Timpson et al., 2014). These studies show differences in the timing of positive and negative deviations (interpreted as population booms and busts) from a null model of a fitted exponential growth. For example, during the first half of the 7th millennium BP, Wessex and Sussex show a negative deviation, whilst other regions such Rhone Languedoc, Rhineland-Hesse, and Paris Basin show a positive deviation (see Fig. 3 in Shennan et al., 2013). For the present study we narrowed our focus to a temporal scope between 8000 and 5000 cal BP, examining a total of 7765 ^{14}C dates from 2268 sites (which constitutes a subset of the database comprising a sample with a ^{14}C age between 8500 and 4500) and using temporal slices of 500 years (6 slices and 5 transitions). Following previous work (Shennan et al., 2013; Timpson et al., 2014) we used a bin size of 200 years ($k = 200$) to reduce the effect of inter-site variability in sampling intensity. We explored various bandwidth values for the spatial kernels; here we illustrate the results for $h = 100$ km which offers a good balance between regional and continental scale of analysis. As for the simulated dataset we obtained our significance levels and false discovery rates using 10,000 simulations; all dates were calibrated using IntCal13 (Reimer et al., 2013) without normalisation.

3. Results

3.1. Simulation study

Fig. 1 shows the SPDRDs of the four regions (Fig. 1-b) and their corresponding growth rate for each of the seven transitions (Fig. 1d). The largest divergence in growth rates can be observed from transition V onwards, when the four regions (and the three trajectories) start to strongly diverge. The results of our spatial analysis (Fig. 2) successfully highlight this pattern, with the highest concentration of hot-spots (higher than expected growth rates) within region C, and cold-spots (lower than expected growth rates) within region B at the transitions V, VII, and VII. More importantly the randomly thinned dataset revealed the same overall pattern, indicating that the method is sufficiently robust to handle uneven sampling intensity. A smaller hot-spots area has also been also identified between regions C and D at transition III in the full dataset, as well as a cold-spot area in region B, again at transition III, in the thinned dataset. Both patterns are also expected from the underlying population from which the radiocarbon dates have been sampled (Fig. 1-a and 1-c).

3.2. Neolithic Europe between 8 k and 5 kBP

Previous work on the same dataset, has shown that at continental scale the SPDRD portrays a general exponential growth, with a “boom” starting from ca. 6000 cal BP, followed by a “bust” at

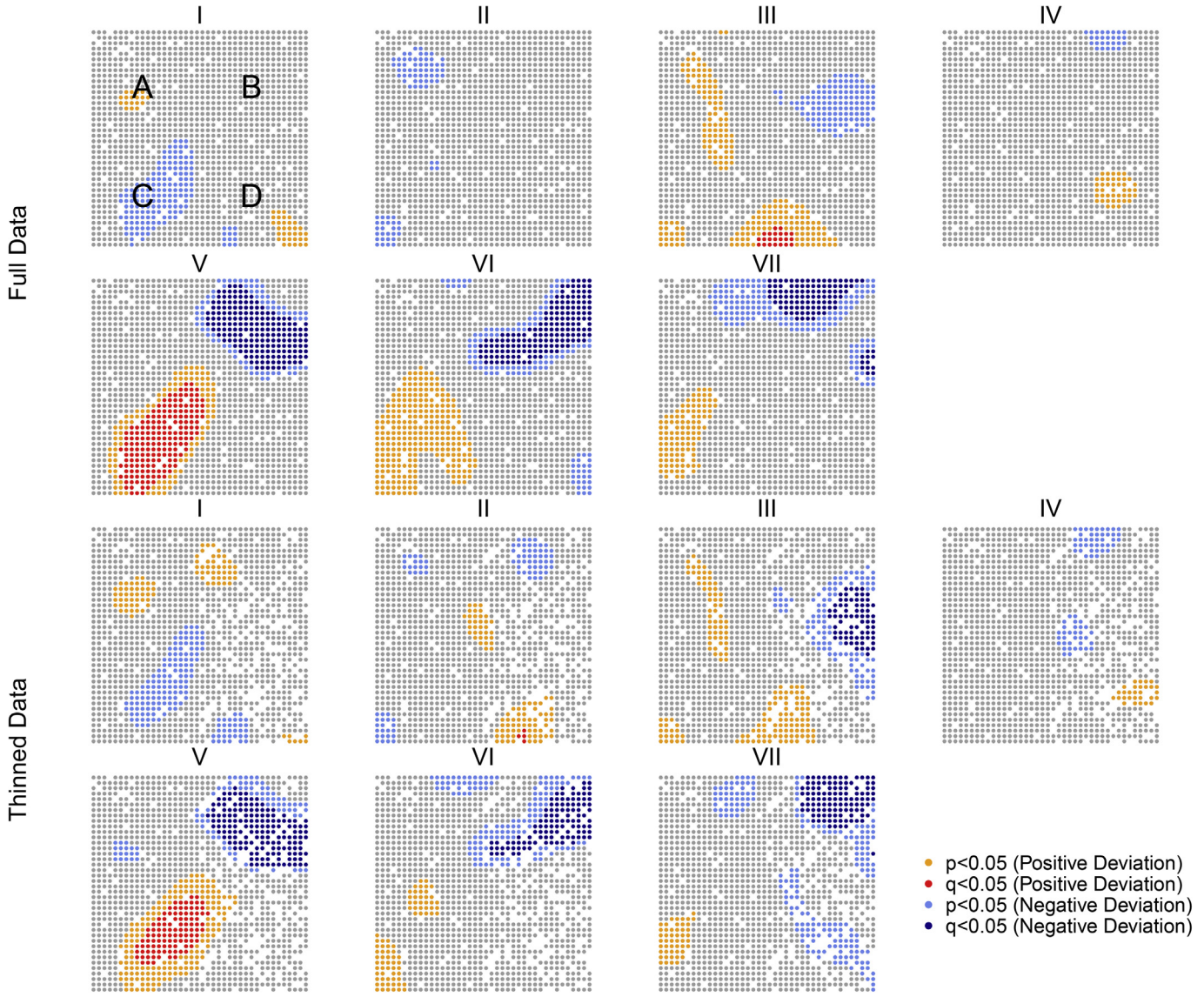

Fig. 2. Results of the spatial permutation test for the full and thinned (uneven sampling intensity across space) datasets.

around 5000 cal. BP (cf. Fig. 2 Shennan et al., 2013). The two events correspond to the transitions IV (from 6.5-6 k BP to 6–5.5 k BP) and V (from 6 to 5.5 k BP to 5.5-5 k BP) in our scheme (Fig. 3), with the latter being the only case where the general growth rate is negative.

Both Shennan et al., 2013 and Timpson et al., 2014 have suggested the presence of divergent population trajectories, with different timings in population boom and busts across Europe. However, because of the differences in sample sizes there was no common standard for comparing these from one region to the next. Fig. 4 shows the pattern, with substantial variation in the local rate of growth across our samples.

The spatial permutation tests show little evidence of local hot/cold spots during the first three transitions (Fig. 5). The few exceptions are all small cold spots: western Ireland in transition I (from 8 to 7.5 k BP to 7.5-7 k BP), Netherlands in transition II (from 7.5-7 k BP to 7–6.5 k BP), corresponding to the effective abandonment of the Low Countries by farming communities at the end of the LBK (Crombé and Vanmontfort, 2007) until the second half of the 7th millennium BP and Central Germany in transition III (from 7 to 6.5 k BP to 6.5-6 k BP; cf. Fig. 3 Shennan et al., 2013, Fig. 3 Timpson et al., 2014). This downturn marks the end of the tradition of Danubian cultures that began in the region with the LBK, prior to the beginning of the southeastern TRB c.5800 BP; the demographic decline is also reflected in the pollen record (Müller, 2001, 92; Zimmermann et al., 2009).

The subsequent transition IV (from 6.5-6 k BP to 6–5.5 k BP) is instead characterised by strong spatial unevenness in growth rates. France, Lowlands, and Moravia are all cold spots, while Britain, Ireland, Denmark, southern Sweden, and Central Germany are all hot spots. The spatial pattern of the hot spots captures in this case the expansion of farming in Britain and in the Baltic area (northern group of the TRB), and a period of renewed growth in Central Germany associated with the construction of a large number of enclosures (Müller, 2001), while the cold spots over large parts of France correspond to a period of stability in the late Chassey complex, which was at its peak at the end of the 7th millennium BP (see contributions to Perrin et al., 2016); it appears as a cold spot because it contrasts so markedly with the overall pattern of growth seen in this period (see Fig. 3a). The transition from the first to the second half of the 6th millennium cal BP (transition V) is

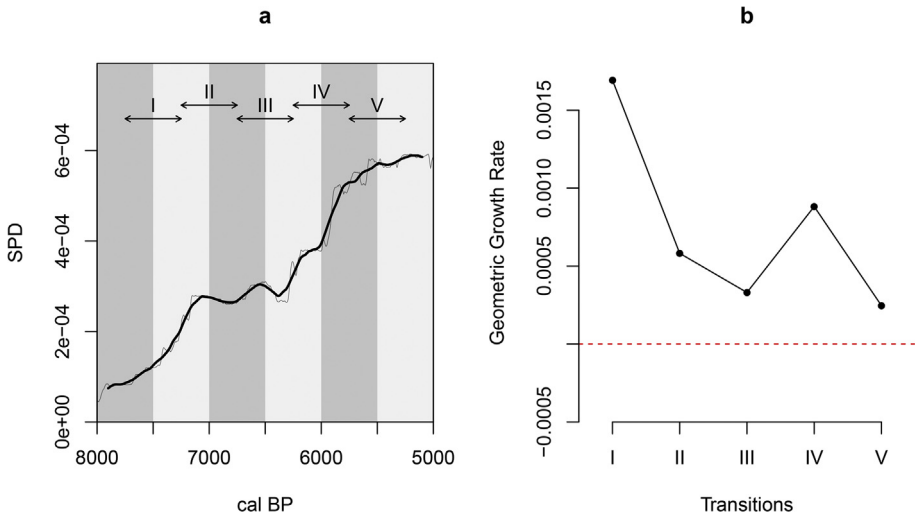

Fig. 3. SPDRD and observed rate of growth for the EUROEVOL dataset.

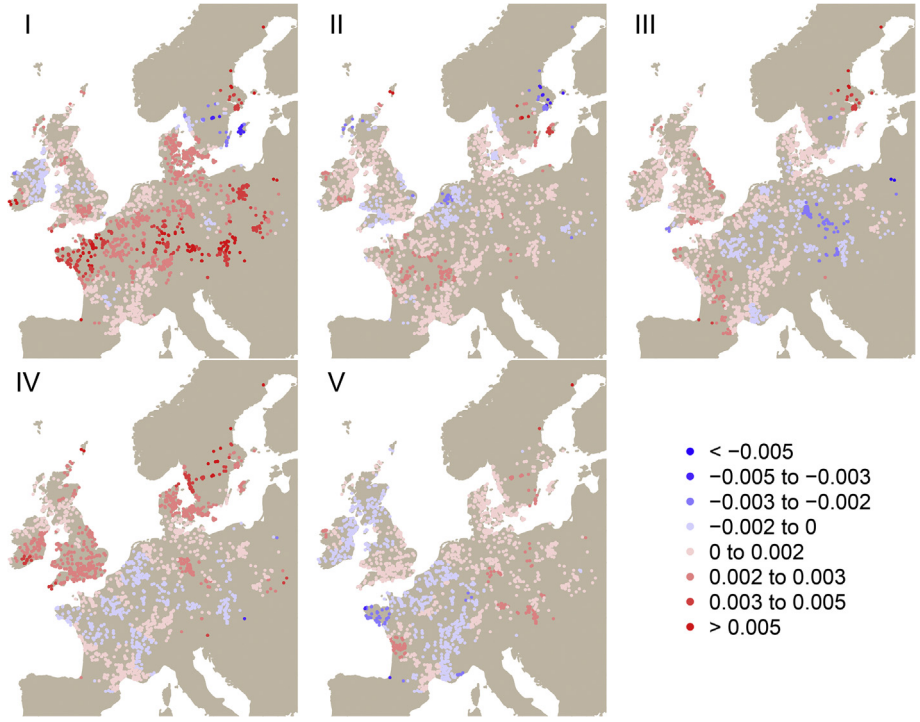

Fig. 4. Local geometric growth rate for each transition in the EUROEVOL dataset (**I**: 8–7.5 k to 7.5–7 k BP; **II**: 7.5-7 k to 7–6.5 k BP; **III**: 7–6.5 k to 6.5-6 k BP; **IV**: 6.5-6 k to 6–5.5 k BP; and **V**: 6–5.5 k to 5.5-5 k BP).

characterised by a highly fragmented picture, with series of highly localised positive and negative deviations from the global trend. This includes hot spots in southwest France, where the Late Neolithic from c.5500 to 5000 BP is characterised by large numbers of multiple ditch enclosures, at least some of which were probably fortified villages (Laporte et al., 2014) and Central Germany, where expansion continues (Müller, 2001); and cold spots in Scotland, Brittany, and Rhineland/Southern Germany; the latter is also identified by Zimmermann et al. (2009) on the basis of the disappearance of settlement nuclei in the region. The evidence from central-eastern Scotland and Brittany is more equivocal. Recent work (Bevan et al. submitted) suggests that in Scotland as a whole population did not decline until the very end of this period, while for Brittany there are no other up-to-date study to make a comparison with this study.

Overall, however, the results suggest that some regions were declining at the same time as others were growing, raising important questions as to why this should be the case and pointing to new directions for research. Without a comparative method of this kind, such questions would not even arise.

4. Discussion

The results of both case studies indicates that the spatial permutation test of the SPDRDs is able to identify instances where the local rate of growth significantly deviates from the general trend

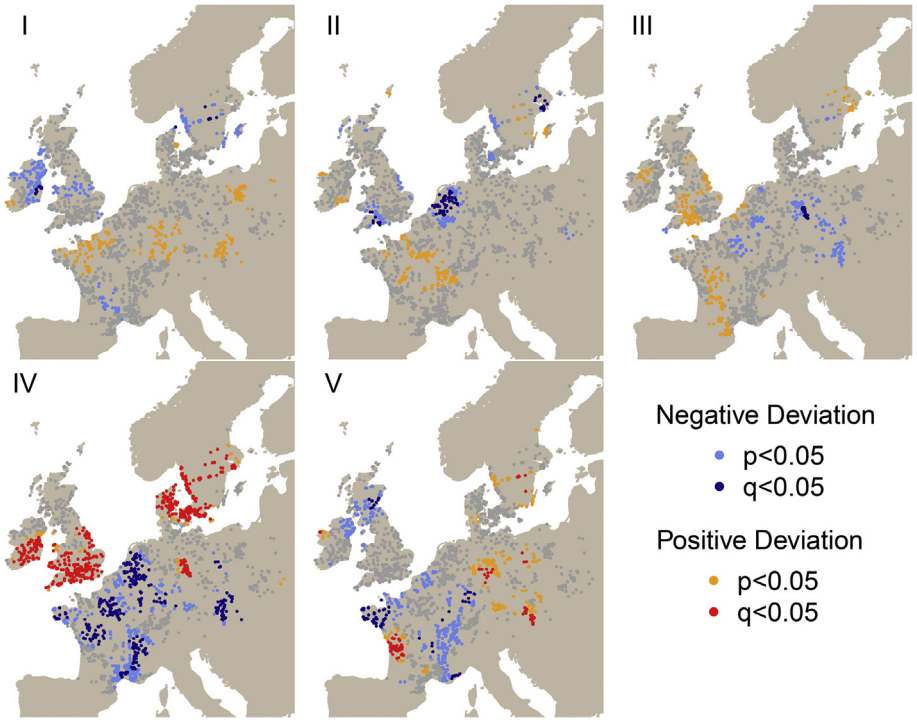

Fig. 5. Spatial permutation test of the EUROEVOL dataset (**I**: 8–7.5 k to 7.5–7 k BP; **II**: 7.5-7 k to 7–6.5 k BP; **III**: 7–6.5 k to 6.5-6 k BP; **IV**: 6.5-6 k to 6–5.5 k BP; and **V**: 6–5.5 k to 5.5-5 k BP).

observed across all regions. The method takes into account differences in the spatially uneven research intensity which characterise most archaeological datasets at regional scales and above, and hence is ideally suited for cross-regional and cross-cultural analysis.

It is worth remembering that the null hypothesis in this case is a spatially homogenous rate of growth across space, and hence strictly speaking the hypothesis should be considered false *a priori*. Hot spots and cold spots are detected in regions where there is a sufficiently large effect size (i.e. a strong deviation from the null) or a high density of sites and ^{14}C dates. Our method is able to reduce type I error via the adoption of *q*-values, but it will inevitably suffer a fairly high level of type II error (i.e. failing to reject a false null hypothesis). Indeed, in our first case study we have no instances of false positives, but we have several transitions and locations with false negatives. This is, to some extent, an inevitable limitation of the proposed method, which is tailored to reduce the kinds of unsupported claims that are typically encountered in spatial and non-spatial analysis of the SPDRDs. To put it in other words, the spatial permutation test proposed here, along with other methods based on Monte Carlo simulation (Timpson et al., 2014) or permutation routines (Crema et al., 2016) should be used as an exploratory tool to detect statistically significant anomalies in the SPDRDs, avoiding simple visual inspections of the data. It is also worth noting that the null hypothesis being evaluated will change as a function of the geographical and temporal extent of the study area, and hence the location of hot-spots and cold-spots might change accordingly. The choice of the temporal and geographical scope of the analysis should hence be carefully justified and appropriately considered in the interpretation of the results. We also acknowledge that the method requires the setting of several free parameters, most notably *k* (the clustering cut-off value for site-level binning of ^{14}C dates) and *h* (the parameter detailing the fall-off of the spatially weighted sum of ^{14}C dates), which should ideally be justified based on archaeological grounds or explored through sensitivity analysis.

Determining whether the hot spots and cold spots detected by the spatial permutation test are genuine episodes of local divergence in past population dynamics, the result of other forms of bias (e.g. a temporally uneven sampling intensity or variations in site-to-population ratio), a mixture of the two, or even processes that are unrelated to past demography is beyond the scope of these analyses. For the EUROEVOL case, independent lines of evidence such as the juvenility index of cemetery data have already confirmed the broad trends depicted by the SPDRDs (Downey et al., 2014). Ultimately the direction to take is to compare multiple proxies of population change and to make a more explicit definition of biases (Davies et al., 2016) and confounding variables (c.f. Kramer-Schadt et al., 2013 in species distribution modelling) either to calibrate (as in Downey et al., 2014) or to test specific hypotheses of demographic changes.

5. Conclusion

The increasing availability of large digital datasets is pushing archaeology into unexplored territories (Bevan, 2015), where comparative and synthetic research can offer new perspectives on key questions about long-term change in human populations (Kintigh et al., 2014). This is an exciting but difficult venture, where data originally collected for a wide range of different purposes are aggregated to answer new questions. The reward is potentially great, but its realisation requires disciplinary investment in the development of methods dedicated to handling the idiosyncratic properties of an archaeological record that has been collected for over 100 years and taking into account its uncertainty; the creation of bespoke techniques is no longer optional.

The analysis of large collections of ^{14}C dates is the prototypical example of this new line of comparative and synthetic research. This paper has introduced a possible solution for dealing with a particular form of research bias — the spatially uneven sampling

intensity of radiocarbon dates — that is typically encountered when we seek to carry out spatio-temporal analysis of radiocarbon dates. Our method, tested on both synthetic and empirical data, can detect instances where local density of radiocarbon dates increased or decreased at a significantly higher or lower growth rate compared to the general pan-regional trend. We note that detecting these divergences is just the first step for a further exploration of the local archaeological record, to establish whether the patterns can be corroborated, in the way briefly illustrated above. Indeed, the comparative and synthetic research we advocate can be truly successful only when its novel insights can drive further research at sub-regional scale, which in turn should feed new exciting lines of enquiry, for example attempting to explain comparative growth trajectories that have to be approached at the macro-scale.

Acknowledgments

The authors are grateful to the constructive comments of three anonymous reviewers.

References

Ames, K.M., 1991. The archaeology of the longue durée: temporal and spatial scale in the evolution of social complexity in southern northwest coast. Antiquity 65, 935–945.
Anselin, L., 1995. Local indicators of spatial association—LISA. Geogr. Anal. 27, 93–115. http://dx.doi.org/10.1111/j.1538-4632.1995.tb00338.x.
Attenbrow, V., Hiscock, P., 2015. Dates and demography: are radiometric dates a robust proxy for long-term prehistoric demographic change? Archaeol. Ocean. 50, 30–36. http://dx.doi.org/10.1002/arco.5052.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society. Ser. B Methodol. 57, 289–300.
Benjamini, Y., Hochberg, Y., 1997. Multiple hypotheses testing with weights. Scand. J. Stat. 24, 407–418. http://dx.doi.org/10.1111/1467-9469.00072.
Bernabeu Aubán, J., García Puchol, O., Barton, M., McClure, S., Pardo Gordó, S., 2016. Radiocarbon dates, climatic events, and social dynamics during the early neolithic in mediterranean Iberia. Quat. Int. 403, 201–210. http://dx.doi.org/10.1016/j.quaint.2015.09.020.
Bevan, A., Crema, E., Li, X., Palmisano, A., 2013. Intensities, interactions and uncertainties: some new approaches to archaeological distributions. In: Bevan, A., Lake, M. (Eds.), Computational Approaches to Archaeological Space. Left Coast Press, Walnut Creek, pp. 27–52.
Bevan, A., 2015. The data deluge. Antiquity 89, 1473–1484. http://dx.doi.org/10.15184/aqy.2015.102.
Brown, W.A., 2017. The past and future of growth rate estimation in demographic temporal frequency analysis: biodemographic interpretability and the ascendance of dynamic growth models. J. Archaeol. Sci. 80, 96–108. http://dx.doi.org/10.1016/j.jas.2017.02.003.
Chaput, M.A., Kriesche, B., Betts, M., Martindale, A., Kulik, R., Schmidt, V., Gajewski, K., 2015. Spatiotemporal distribution of Holocene populations in north America. PNAS 112, 12127–12132. http://dx.doi.org/10.1073/pnas.1505657112.
Chaput, M.A., Gajewski, K., 2016. Radiocarbon dates as estimates of ancient human population size. Anthropocene, AAG Hum-Induce Environ. Chg. 15, 3–12. http://dx.doi.org/10.1016/j.ancene.2015.10.002.
Collard, M., Edinborough, K., Shennan, S., Thomas, M.G., 2010. Radiocarbon evidence indicates that migrants introduced farming to Britain. J. Archaeol. Sci. 37, 866–870.
Contreras, D.A., Meadows, J., 2014. Summed radiocarbon calibrations as a population proxy: a critical evaluation using a realistic simulation approach. J. Archaeol. Sci. 52, 591–608.
Crema, E.R., Bianchi, E., 2013. Looking for patterns in the noise: non-site spatial-analysis. In: Mulazzani, S. (Ed.), Le Capsien De Hergla (Tunisie).Culture, Environnement Et Économie, pp. 385–395.
Crema, E.R., Habu, J., Kobayashi, K., Madella, M., 2016. Summed probability distribution of 14 C dates suggests regional divergences in the population dynamics of the Jomon period in eastern Japan. PLos One 11, e0154809. http://dx.doi.org/10.1371/journal.pone.0154809.
Crombé, P., Vanmontfort, B., 2007. The neolithisation of the Scheldt basin in western Belgium. Proc. Br. Acad. 144, 263–285.
Davies, B., Holdaway, S.J., Fanning, P.C., 2016. Modelling the palimpsest: an exploratory agent-based model of surface archaeological deposit formation in a fluvial arid Australian landscape. Holocene 26, 450–463. http://dx.doi.org/10.1177/0959683615609754.
Downey, S.S., Bocaege, E., Kerig, T., Edinborough, K., Shennan, S., 2014. The neolithic demographic transition in Europe: correlation with juvenility index supports interpretation of the summed calibrated radiocarbon date probability distribution (SCDPD) as a valid demographic proxy. PLoS One 9, e105730. http://dx.doi.org/10.1371/journal.pone.0105730.
Eve, S.J., Crema, E.R., 2014. A house with a view? Multi-model inference, visibility fields, and point process analysis of a Bronze Age settlement on Leskernick Hill (Cornwall, UK). J. Archaeol. Sci. 43, 267–277.
Fitzpatrick, A., 1987. The structure of a distribution map: problems of sample bias and quantitative studies. In: Tomasevic-Buck, T., Kellner, H. (Eds.), Rei Cretariae Romanae Fautorum Acta. Rei Cretariae Romanae Fautores. London and Oxford, pp. 79–112.
Fotheringham, S.A., Brunsdon, C., Charlton, M., 2000. Quantitative Geography: Perspectives on Spatial Data Analysis. Sage Publications, London.
Gamble, C., Davies, W., Pettitt, P., Hazelwood, L., Richards, M., 2005. The archaeological and genetic foundations of the european population during the late glacial: implications for. Camb. Archaeol. J. 15, 193–223. http://dx.doi.org/10.1017/S0959774305000107.
Gajewski, K., Munoz, S., Peros, M., Viau, A., Morlan, R., Betts, M., 2011. The canadian archaeological radiocarbon database (CARD): archaeological 14C dates in north America and their paleoenvironmental context. Radiocarbon 53, 371–394. http://dx.doi.org/10.2458/azu_js_rc.53.3470.
García Puchol, O., Diez Castillo, A., Pardo-Gordó, S., 2017. New insights into the neolithisation process in southwest Europe according to spatial density analysis from calibrated radiocarbon dates. Archaeol. Anthropol. Sci. http://dx.doi.org/10.1007/s12520-017-0498-1.
Gayo, E.M., Latorre, C., Santoro, C.M., 2015. Timing of occupation and regional settlement patterns revealed by time-series analyses of an archaeological radiocarbon database for the South-Central Andes (16°–25°S). quaternary international. Palaeodemogr. South. S. Am. 356, 4–14. http://dx.doi.org/10.1016/j.quaint.2014.09.076.
Getis, A., Ord, J.K., 1996. Local spatial statistics: an overview. In: Longley, P., Batty, M. (Eds.), Spatial Analysis: Modelling in a GIS Environment. GeoInformation International, Cambridge, pp. 269–285.
Goldberg, A., Mychajliw, A.M., Hadly, E.A., 2016. Post-invasion demography of prehistoric humans in South America. Nature 532, 232–235. http://dx.doi.org/10.1038/nature17176.
Grove, M., 2011. A spatio-temporal kernel method for mapping change in prehistoric land-use patterns. Archaeometry 53, 1012–1030.
Hiscock, P., Attenbrow, V., 2016. Dates and demography? The need for caution in using radiometric dates as a robust proxy for prehistoric population change. Archaeol. Ocean. 51, 218–219. http://dx.doi.org/10.1002/arco.5096.
Hodder, I., Orton, C., 1976. Spatial Analysis in Archaeology. Cambridge University Press, Cambridge.
Kelly, R.L., Surovell, T.A., Shuman, B.N., Smith, G.M., 2013. A continuous climatic impact on Holocene human population in the Rocky Mountains. Proc. Natl. Acad. Sci. 110, 443–447.
Kintigh, K.W., Altschul, J.H., Beaudry, M.C., Drennan, R.D., Kinzig, A.P., Kohler, T.A., Limp, W.F., Maschner, H.D.G., Michener, W.K., Pauketat, T.R., Peregrine, P., Sabloff, J.A., Wilkinson, T.J., Wright, H.T., Zeder, M.A., 2014. Grand challenges for archaeology. Am. Antiq. 79, 5–24. http://dx.doi.org/10.7183/0002-7316.79.1.5.
Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schröder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A.K., Augeri, D.M., Cheyne, S.M., Hearn, A.J., Ross, J., Macdonald, D.W., Mathai, J., Eaton, J., Marshall, A.J., Semiadi, G., Rustam, R., Bernard, H., Alfred, R., Samejima, H., Duckworth, J.W., Breitenmoser-Wuersten, C., Belant, J.L., Hofer, H., Wilting, A., 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. Divers. Distrib. 19, 1366–1379. http://dx.doi.org/10.1111/ddi.12096.
Laporte, L., Bizien-Jaglin, C., Guyodo, J.N., 2014. Enceintes neolithiques de l'Ouest de la France: une archeologie des fosses? In: Joussaume, R., Large, J.M. (Eds.), Enceintes néolithiques de l'Ouest de la France: de la Seine à la Gironde. Association des Publications Chauvinoises, Chauvigny, pp. 455–488.
Manning, K., Timpson, A., 2014. The demographic response to Holocene climate change in the Sahara. Quat. Sci. Rev. 101, 28–35. http://dx.doi.org/10.1016/j.quascirev.2014.07.003.
Manning, K., Colledge, S., Crema, E., Shennan, S., Timpson, A., 2016. The cultural evolution of neolithic Europe. EUROEVOL dataset 1: sites, phases and radiocarbon data. J. Open Archaeol. Data 5. http://dx.doi.org/10.5334/joad.40.
Miller, D.S., Gingerich, J.A.M., 2013. Regional variation in the terminal Pleistocene and early Holocene radiocarbon record of eastern North America. Quat. Res. 79, 175–188. http://dx.doi.org/10.1016/j.yqres.2012.12.003.
Mökkönen, T., 2014. Archaeological radiocarbon dates as a population proxy: a skeptical view. Fennosc. Archaeol. 31, 125–134.
Müller, J., 2001. Soziochronologische Studien zum Jung- und Spätneolithikum im Mittelelbe-Saale-Gebiet (4100-2700 v. Chr.). Verlag Marie Leidorf, Rahden.
North, B.V., Curtis, D., Sham, P.C., 2002. A note on the calculation of empirical P values from Monte Carlo procedures. Am. J. Hum. Genet. 71, 439–441.
Onkamo, P., Kammonen, J., Pesonen, P., Sundell, T., Moltchanova, E., Oinonen, M., Haimila, M., Arjas, E., 2012. Bayesian spatiotemporal analysis of radiocarbon dates from eastern fennoscandia. Radiocarbon 54, 649–659. http://dx.doi.org/10.2458/azu_js_rc.54.16136.
Perez, S.I., Postillone, M.B., Rindel, D., Gobbo, D., Gonzalez, P.N., Bernal, V., 2016. Peopling time, spatial occupation and demography of Late Pleistocene–Holocene human population from Patagonia. Quat. Int. 425, 214–223. http://dx.doi.org/10.1016/j.quaint.2016.05.004.
Perrin, T., Chambon, P., Gibaja, J.F., Goude, G. (Eds.), 2016. Le Chasséen, des Chasséens … Retour sur une culture nationale et ses parallèles. Archives

d'Écologie Préhistorique, Toulouse. Sepulcres de fossa, Cortaillod, Lagozza.
Porčić, M., Nikolić, M., 2016. The Approximate Bayesian Computation approach to reconstructing population dynamics and size from settlement data: demography of the mesolithic-neolithic transition at Lepenski Vir. Archaeol. Anthropol. Sci. 8, 169–186. http://dx.doi.org/10.1007/s12520-014-0223-2.
Premo, L., 2004. Local spatial autocorrelation statistics quantify multi-scale patterns in distributional data: an example from the Maya Lowlands. J. Archaeol. Sci. 31, 855–866.
R Core Team, 2016. R: A Language and Environment for Statistical Computing. URL. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Reimer, P.J., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Ramsey, C.B., Buck, C.E., Cheng, H., Edwards, R.L., Friedrich, M., Grootes, P.M., Guilderson, T.P., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T.J., Hoffmann, D.L., Hogg, A.G., Hughen, K.A., Kaiser, K.F., Kromer, B., Manning, S.W., Niu, M., Reimer, R.W., Richards, D.A., Scott, E.M., Southon, J.R., Staff, R.A., Turney, C.S.M., van der, Plicht, J., 2013. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 Years cal BP. Radiocarbon 55, 1869–1887. http://dx.doi.org/10.2458/azu_js_rc.55.16947.
Rick, J.W., 1987. Dates as Data: an examination of the Peruvian radiocarbon record. Am. Antiq. 1987 (52), 55–73.
Smith, M., 2016. The use of summed-probability plots of radiocarbon data in archaeology. Archaeol. Ocean. 51, 214–215. http://dx.doi.org/10.1002/arco.5094.
Stolar, J., Nielsen, S.E., 2015. Accounting for spatially biased sampling effort in presence-only species distribution modelling. Divers. Distrib. 21, 595–608. http://dx.doi.org/10.1111/ddi.12279.
Syfert, M.M., Smith, M.J., Coomes, D.A., 2013. The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. PLos One 8, e55158. http://dx.doi.org/10.1371/journal.pone.0055158.
Shennan, S., Edinborough, K., 2007. Prehistoric population history: from the late glacial to the late neolithic in central and northern Europe. J. Archaeol. Sci. 34, 1339–1345.
Shennan, S., Downey, S.S., Timpson, A., Edinborough, K., Colledge, S., Kerig, T., Manning, K., Thomas, M.G., 2013. Regional population collapse followed initial agriculture booms in mid-Holocene Europe. Nat. Commun. 4, ncomms3486 http://dx.doi.org/10.1038/ncomms3486.
Surovell, T.A., Brantingham, P.J., 2007. A note on the use of temporal frequency distributions in studies of prehistoric demography. J. Archaeol. Sci. 34, 1868–1877.
Surovell, T.A., Finley, J.B., Smith, G.M., Brantingham, P.J., Kelly, R., 2009. Correcting temporal frequency distributions for taphonomic bias. J. Archaeol. Sci. 36, 1715–1724.
Tallavaara, M., Pesonen, P., Oinonen, M., Seppä, H., 2014. The mere possibility of biases does not invalidate archaeological population proxies - response to teemu mökkönen. Fennosc. Archaeol. 31, 135–140.
Timpson, A., Colledge, S., Crema, E., Edinborough, K., Kerig, T., Manning, K., Thomas, M.G., Shennan, S., 2014. Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: a new case-study using an improved method. J. Archaeol. Sci. 52, 549–557. http://dx.doi.org/10.1016/j.jas.2014.08.011.
Timpson, A., Manning, K., Shennan, S., 2015. Inferential mistakes in population proxies: a response to Torfing's "Neolithic population and summed probability distribution of 14C-dates." J. Archaeol. Sci. 63, 199–202.
Torfing, T., 2015. Neolithic population and summed probability distribution of 14C-dates. J. Archaeol. Sci. 63, 193–198.
Wang, C., Lu, H., Zhang, J., Gu, Z., He, K., 2014. Prehistoric demographic fluctuations in China inferred from radiocarbon data and their linkage with climate change over the past 50,000 years. Quat. Sci. Rev. 98, 45–59. http://dx.doi.org/10.1016/j.quascirev.2014.05.015.
Weninger, B., Clare, L., Jöris, O., Jung, R., Edinborough, K., 2015. Quantum theory of radiocarbon calibration. World Archaeol. 47, 543–566. http://dx.doi.org/10.1080/00438243.2015.1064022.
Williams, A.N., Ulm, S., Smith, M., Reid, J., 2014. A database of 14C and Non-14C ages from archaeological sites in Australia - composition, compilation and review (data paper). Internet Archaeol. 36 http://dx.doi.org/10.11141/ia.36.6.
Williams, A.N., Ulm, S., 2016. Radiometric dates are a robust proxy for long-term demographic change: a comment on Attenbrow and Hiscock (2015). Archaeol. Ocean. 51, 216–217. http://dx.doi.org/10.1002/arco.5095.
Zahid, H.J., Robinson, E., Kelly, R.L., 2016. Agriculture, population growth, and statistical analysis of the radiocarbon record. Proc. Natl. Acad. Sci. 113, 931–935.
Zimmermann, A., Wendt, K.P., Frank, T., Hilpert, J., 2009. Landscape archaeology in central Europe. Proc. Prehist. Soc. 75, 1–53.