



# Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: a new case-study using an improved method



Adrian Timpson<sup>a, b, \*</sup>, Sue Colledge<sup>a</sup>, Enrico Crema<sup>a</sup>, Kevan Edinborough<sup>a</sup>, Tim Kerig<sup>c</sup>, Katie Manning<sup>a</sup>, Mark G. Thomas<sup>b</sup>, Stephen Shennan<sup>a</sup>

<sup>a</sup> Institute of Archaeology, University College London, 31-34 Gordon Square, London WC1H 0PY, UK

<sup>b</sup> Research Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK

<sup>c</sup> Archäologisches Institut, Albertus-Magnus-Platz, University of Cologne, D-50923 Köln, Germany

## ARTICLE INFO

### Article history:

Available online 2 September 2014

### Keywords:

Neolithic  
Radiocarbon  
Demography  
Simulation  
Europe  
Monte Carlo  
SPD  
Summed probability distribution

## ABSTRACT

In a previous study we presented a new method that used summed probability distributions (SPD) of radiocarbon dates as a proxy for population levels, and Monte-Carlo simulation to test the significance of the observed fluctuations in the context of uncertainty in the calibration curve and archaeological sampling. The method allowed us to identify periods of significant short-term population change, caveated with the fact that around 5% of these periods were false positives. In this study we present an improvement to the method by applying a criterion to remove these false positives from both the simulated and observed distributions, resulting in a substantial improvement to both its sensitivity and specificity. We also demonstrate that the method is extremely robust in the face of small sample sizes. Finally we apply this improved method to radiocarbon datasets from 12 European regions, covering the period 8000–4000 BP. As in our previous study, the results reveal a boom-bust pattern for most regions, with population levels rising rapidly after the local arrival of farming, followed by a crash to levels much lower than the peak. The prevalence of this phenomenon, combined with the dissimilarity and lack of synchronicity in the general shapes of the regional SPDs, supports the hypothesis of endogenous causes.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

## 1. Introduction

Population size and density are key variables in human evolution. They represent important outcomes of evolutionary adaptation, and have strong feedback relationships with key processes such as: the transmission, selection and drift of both genetic and cultural information; infectious disease dynamics; land and resource use; niche construction; economic cycles and sustainability. To understand human evolution it is therefore necessary to estimate regional population fluctuations, and to identify their causes and consequences. Major advances are now being made in this field due to the growing availability of modern and ancient genetic data and associated modelling approaches (e.g. Li and

Durbin, 2011). However, estimates of population size from these data generally lack adequate chronological and/or spatial resolution, or the data are too few in number, to draw meaningful inferences about their relationship with these key processes.

Directly dated archaeological site information does not suffer from these problems but, with some recent exceptions (e.g. Bocquet-Appel (2002) using cemetery age distributions, Zimmermann et al. (2004) using site spatial distributions, and Hinz et al. (2012) using summed radiocarbon probabilities), archaeologists, in Europe at least, have been strikingly reluctant to make demographic inferences from such data, and are generally keener to emphasise the pitfalls than the possibilities. When Rick (1987) proposed using summed date distributions as data for the purpose of reconstructing spatial-temporal variation in coastal-highland settlement practices during the Peruvian preceramic period, an important new weapon was added to the archaeologist's armoury. In his inferential chain, Rick laid out three main assumptions that underpin this approach; firstly that more dateable objects will be deposited during periods when the population was

\* Corresponding author. Research Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK. Tel.: +44 2076794397.

E-mail address: [a.timpson@ucl.ac.uk](mailto:a.timpson@ucl.ac.uk) (A. Timpson).

larger, secondly that more deposits will lead to more objects preserved in the archaeological record, and thirdly that more preserved objects will lead to more dateable material eventually recovered by archaeologists. Joining these together gives us the assumption of a monotonic relationship between the population size and the amount of radiocarbon dates recovered (Collard et al. 2010). Therefore a suitable radiocarbon database can be used to construct a time-series by summing each date's probability distribution, and the fluctuations in this time-series can then be used as a proxy for changing population size.

Of course the extent to which these assumptions are satisfied can be difficult to determine. The law of large numbers predicts that larger sample sizes should more fairly represent the archaeological record, but this may already be a taphonomically biased representation of the original deposits. Some control can be achieved by using radiocarbon dates from a confined spatial region, small enough for taphonomic losses to be considered spatially homogeneous. However, this necessarily reduces sample sizes, and so a balance must be found. Even in this simple case, where the analysis deals only with a local pattern, we can expect constant homogeneous taphonomic losses to manifest as a gradual loss over time in the archaeological record, and therefore a long-term exponential increase in the summed distribution.

Whilst the utility of this approach is reflected in its increasing application, the biases and assumptions noted in Rick's chain of inference have also been subject to increasing critical scrutiny (Rick, 1987 Fig. 1; Surovell et al. 2009; Williams, 2012; Bamforth and Grund, 2012). Three major issues that persist are; the impact of sample size, fluctuations in the radiocarbon calibration curve – which have the effect of concentrating dates in some time periods and spreading them out across others – and the effect of differential taphonomic and archaeological recovery processes on what is available for dating.

In our previous study (Shennan et al. 2013) we have shown that many of the problems and biases raised by the standard approach of summing radiocarbon dates can be resolved. Despite this, criticisms persist; most recently for example, Contreras and Meadows (2014) again raise these concerns. The authors simulate a radiocarbon dataset by sampling from a prior 'assumed true' population curve (using Bennett's population reconstruction of the European Black Death AD 1000–1700, and McCaa's population reconstruction of Central Mexico AD 1000–1800), and then comment on the dissimilarity between the sampled summed probability distribution and the 'true' population curve from which it was sampled. In principle this is a sensible approach, which should be expected to demonstrate good congruence as the sample size increases; however the authors argue the contrary, that there is poor congruence, and conclude the method is unreliable.

There is a simple explanation for this. Because of the interference effect of wiggles in the calibration curve, spurious fluctuations exist on a scale below c.200 years, rendering this method quite useless for any time-series shorter than a few thousand years. This is simply a matter of analysing at the appropriate scale – the effect of these wiggles is invisible and irrelevant at the scale of tens of thousand years. As with our previous study, we apply this method to dates spanning several thousand years, before trimming the summed distribution down to a 4000 year period of interest, to avoid edge effects. Furthermore we plot a 200 year rolling mean, to discourage the reader from over-interpreting smaller scale features. In contrast, Contreras and Meadows invoke a straw man by simulating dates over the inappropriately short time ranges of 700 years and 800 years, so that the shape of their distribution is dominated by these spurious short-term wiggles. They obfuscate matters further by plotting the simulated distribution over a wider 1200 year range, so as to include yet more spurious edge effects outside the range covered by the sampled data.

Shennan et al. (2013) also showed that a more comprehensive Monte-Carlo simulation-based method, which generates simulated date distributions under a fitted null model, can be used to test features in the observed dataset for statistically significant patterns. The results of this Monte-Carlo Summed Probability Distribution method (MCSPD-method) can be supplemented by comparing the radiocarbon population proxy with other proxies, based on independent evidence and different assumptions. Thus, Woodbridge et al. (in press) compared this population proxy for Britain with independent evidence for forest clearance, based on pollen analysis, which serves as an indicator of human environmental impact and hence population size, and found a strong correlation: peaks in the summed date distribution correspond to more open environments and troughs to more extensive forest cover. Other studies of the European Neolithic have produced the same result (see Hinz et al. 2012; Whitehouse et al. in press; Lechterbeck et al. in press).

Shennan et al. (2013) addressed the question of whether the arrival of farming in the different regions of Europe was associated with a significant departure from a fitted null-model of long-term exponential growth that characterises both global population history (e.g. McEvedy and Jones, 1978) and the increased survival of the archaeological record towards the present. Results for the majority of the European regions showed significant departures from this null model, and indicate that boom and bust fluctuations followed the arrival of farming. The occurrence of population booms – periods of rapid population growth – associated with the local arrival of farming was unsurprising, on the basis of both theory (Ammerman and Cavalli-Sforza, 1971) and inferences of increased growth rates derived from cemetery age-at-death distributions (Bocquet-Appel, 2002). However, the consistent evidence for population 'busts' contradicts standard views about the long-term impact of agriculture on population levels. Furthermore, cross-correlation of the population fluctuations with climate data did not support the hypothesis that the fluctuations were climate-driven.

This paper pursues a similar agenda by examining dates from another twelve European regions (see map Fig. 1, and Appendix 2 for date sources) to see if they continue to support the boom-bust pattern, but does so by means of an improvement to the existing MCSPD-method that was presented in Shennan et al. (2013). We provide a detailed description of the improved method, and demonstrate its power using one of the twelve regions as a test set, by progressively sampling smaller and smaller training datasets and comparing the results. Finally, we examine the population reconstructions for the twelve regions and discuss their implications.

## 2. Data

As with our previous study, radiocarbon dates for each study area were selected from the EUROEVOL project database. Once again we used a fully inclusive approach on the basis that inaccurate dates would obscure any genuine underlying patterns, thus having a conservative effect, and that the larger the sample, the closer it will approximate the true distribution (see Shennan et al. 2013 for details).

## 3. Improvement to computational method

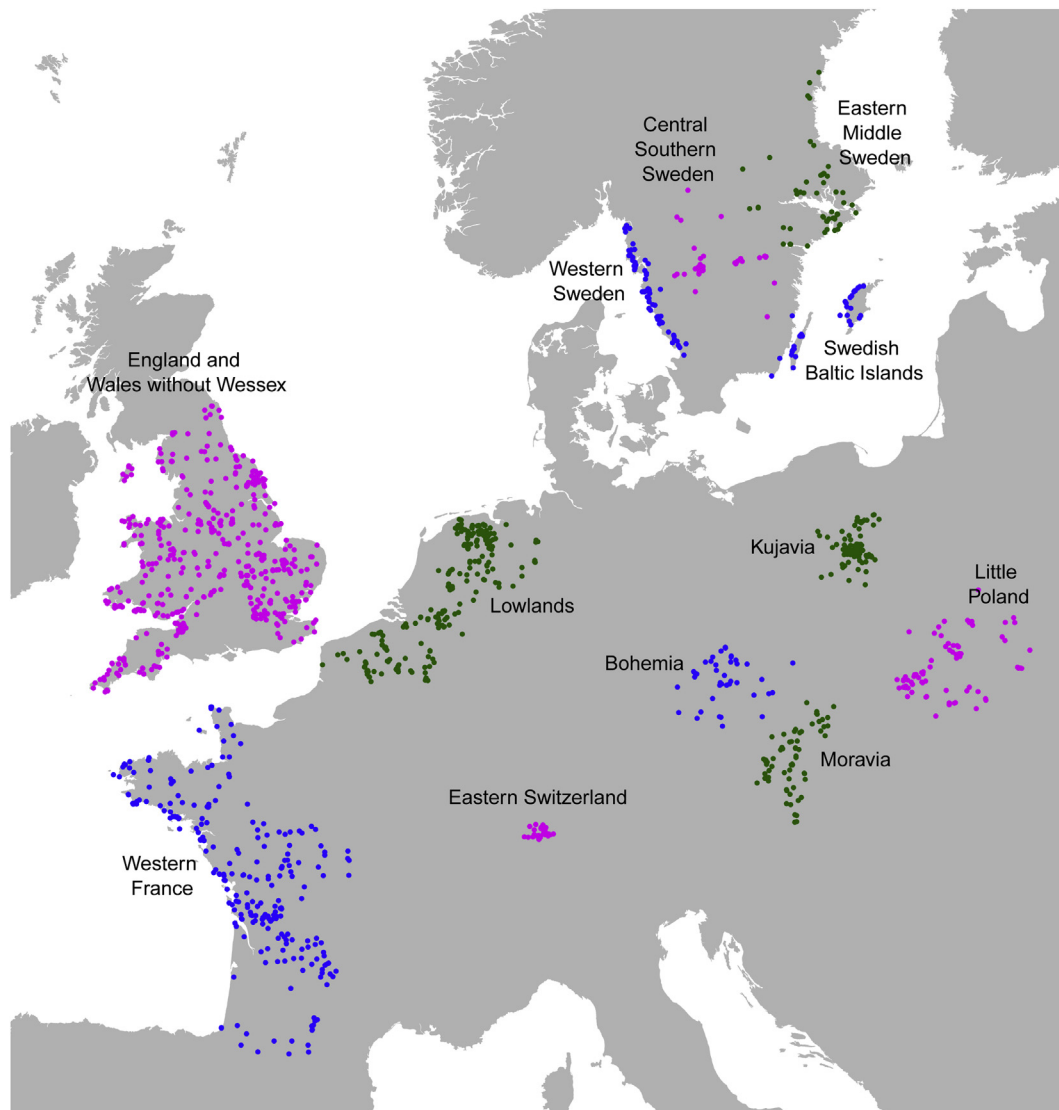
By definition, approximately 5% of any SPD (constructed from observed data or simulated data) will be falsely considered unusually high/low density by the existing method, and reported as locally significant (highlighted in red/blue respectively in the figures below). This is because 5% of any random data falls outside its 95% confidence interval, and can be loosely considered as 'false positive'

points on the SPD plot, in so far as all the red/blue regions can be considered ‘positive’ points. A ‘global’  $p$ -value informs us if overall there is a significant departure from the null for the entire time series, since this  $p$ -value is estimated by comparing a single global summary statistic from the observed SPD with a distribution of the same statistic from all the simulated SPDs. However, we are left with difficulties in interpreting precisely which time intervals truly depart from expectation under the null-model, and which are ‘false’. Here we introduce an additional function that seeks to filter out the points that are most likely to be ‘false positives’ from both the observed SPD and each simulated SPD (for a flowchart see Fig. 4; details of the entire method are fully explained in Appendix 1; ‘details of computational method’). The new ‘false positive remover’ function uses the principle that the ‘false positive’ points were not caused by an interesting underlying signal in the data, and instead are randomly distributed through a SPD. Therefore the ‘false positives’ are more likely to occur alone than in pairs, more likely in pairs than triplets etc. In contrast, ‘true positives’ in an observed SPD are more likely to have low entropy, i.e. exhibit order, and therefore occur in blocks, since they are caused by some underlying population process. The new function therefore filters out single positive points on an SPD, followed by points with only one neighbouring

point, until a maximum of 5% of the SPD has been removed from the category of positive. The immediate affect of this is to selectively reduce the amount of red/blue in a manner that tends to preferentially reduce the false positive points in the observed SPD, thus improving the specificity of the plots. However, the new function is also applied equally to the simulated SPDs, which has a very different effect, since the simulated SPDs are used to calculate the global  $p$ -value. The simulated SPDs are assigned fewer positive points (points outside the 95% CI), which lowers the summary statistic threshold used in calculating the global  $p$ -value. This results in the test having increased sensitivity, since it is now better at successfully rejecting an incorrect null. Overall these two effects combine to substantially increase the power and usefulness of the method.

#### 4. Demonstrating the power of the MCSPD-method

We use one of the twelve regional datasets – Eastern Middle Sweden (EMS) to test the efficacy and statistical power of the improved MCSPD-method, which includes the addition of the ‘false positive remover’ function (Fig. 2). The EMS dataset was selected as a suitable candidate for this demonstration since it exhibits a



**Fig. 1.** Showing the twelve regions tested. Dots represent the site locations. Colours are merely to aid visual differentiation where regions are close to each other. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

simple and coherent pattern, which despite the small sample size (samples  $N = 93$ , bins  $N = 63$ ) corresponds well to that seen in a much larger sample of dates from the same region (Hinz et al., 2012, Fig. 3;  $N$  samples = 243).

Initially the full (100%) EMS dataset was analysed using all 93 samples. This was repeated for 7 further tests, with each subsequent subset being one third smaller than the previous subset, but always randomly sampled from the original dataset. The rationale behind this approach is that by assessing the similarity of the SPDs (and the short term periods of significance) between subsets, we can estimate the minimum sample size required to recover a shape that is fairly representative of the 100% SPD. This also provides a way of critically assessing the  $p$ -values generated by the method, since both the  $p$ -value for each subset, and the similarity of shape between subsets, are indicators that the shape is not a random artefact of a small sample.

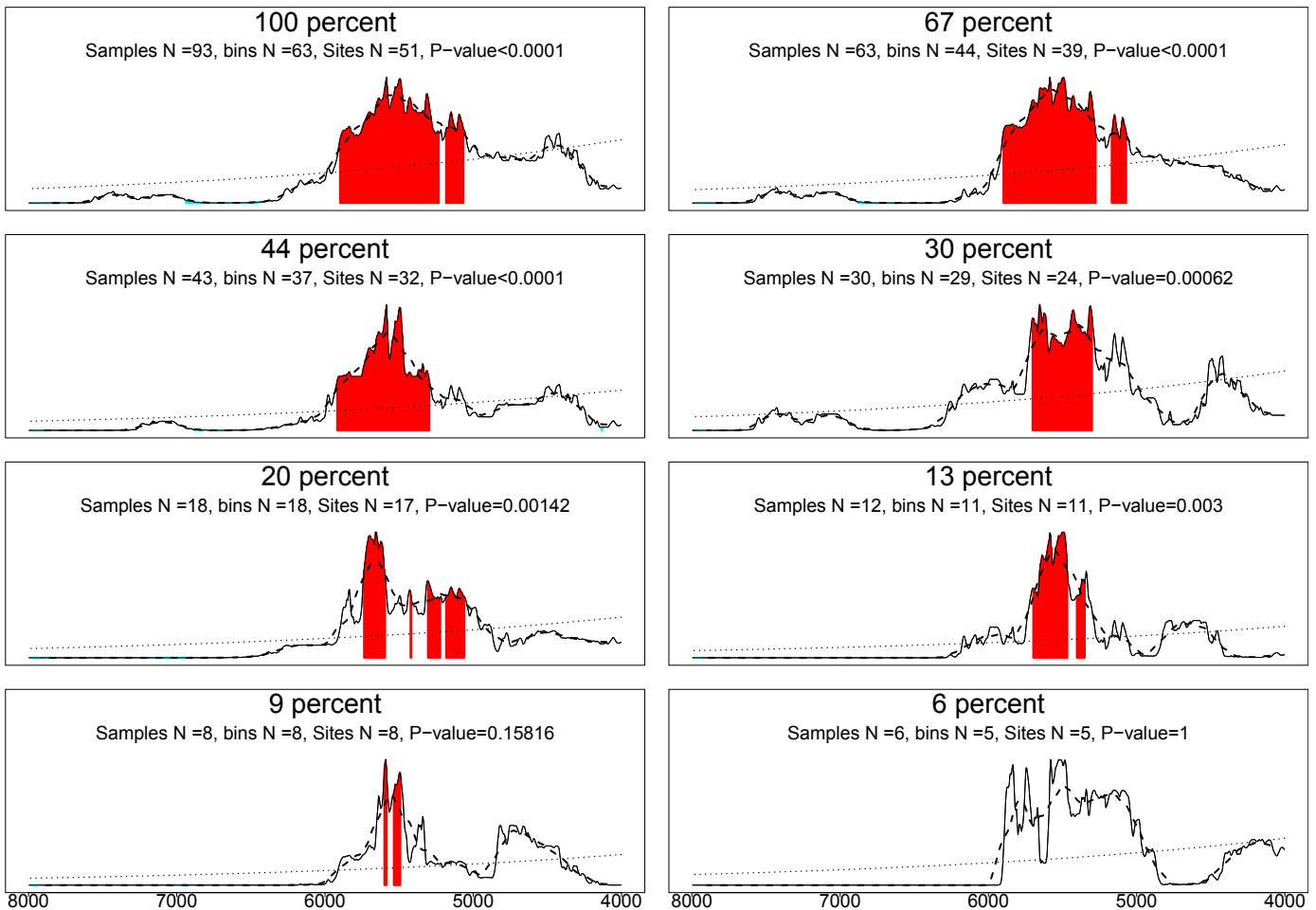
This demonstration utilises the consequences of the law of large numbers, which predicts that as sample sizes increase, the sample distribution approaches the true distribution; therefore as sample sizes decrease, the shapes of the sample distributions will become increasingly different.

The results show remarkable similarity in the broad scale shape of all SPDs, even at a sample size of 6% of the full dataset, comprising just six  $^{14}\text{C}$  dates across 4000 years (contra Williams, 2012: 587). Local regions of high density (red) also exhibit good

synchronicity across all random subsets. When a strong pattern of clustering is present in the data the global  $p$ -values suggest that the improved method has the statistical power to detect significance with sample sizes as small as 12 dates across 4000 years (test using 13% of full EMS dataset yielded a  $p$ -value of 0.003). At this level the similarity with the full dataset is remarkable, both in terms of the shape of the SPD and the local regions of significance (red).

Regions of low density (blue) are sparse in the full dataset and disappear entirely on subsequent smaller subsets. This phenomenon is driven by the small sample size (compare with larger sections of blue from the more substantial datasets of 'England and Wales without Wessex' and 'Western France', Fig. 3) and is a consequence of the fact that absence of evidence is not evidence of absence in a small dataset (in a very small dataset the lower bound of the confidence interval falls to zero, but it is not possible for any part of an SPD to be smaller than zero), whilst as sample sizes increase the contrary becomes true — the absence of dates (low density) increasingly becomes evidence of absence.

In contrast to the results from Fig. 2, Fig. 3 shows that the same method applied to a similar sample size for Swedish Baltic Islands revealed no significance. This might seem counterintuitive since there are clear fluctuations in the SPD, and seems a good candidate for subjective disagreement about the extent to which those fluctuations are significant, or merely the random effects of sampling.



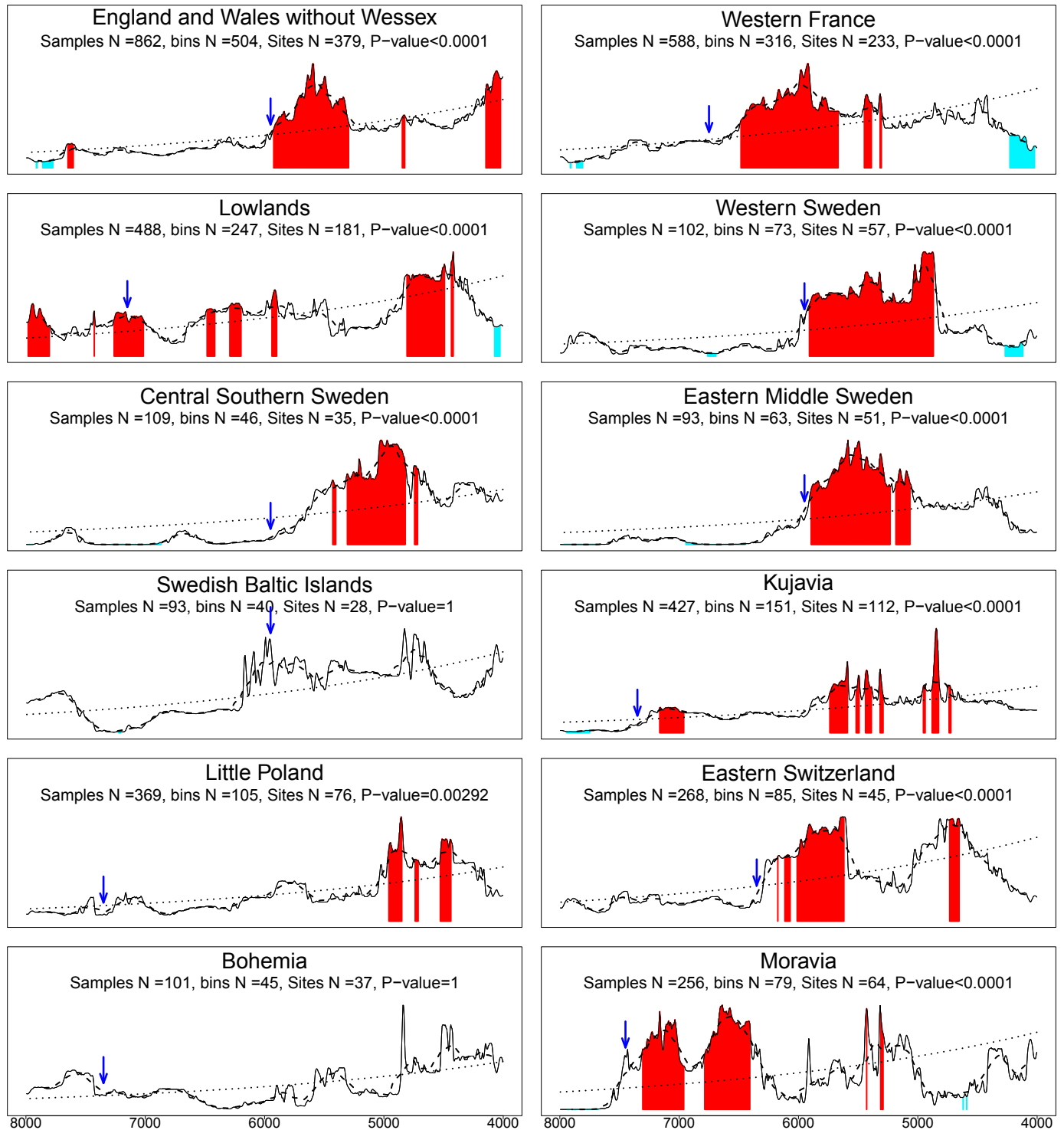
**Fig. 2.** Showing the effects of progressively smaller subsets of the Eastern Middle Sweden (EMS) dataset. Black line shows the summed probability distribution (SPD); dotted line shows a rolling 200 year mean of the SPD; dashed line shows the fitted null exponential model (for consistency we used the same null model as in the previous study, see section 5 below); red sections are local regions of unusually high density, blue shows regions of unusually low density. At 13% (12 samples)  $p$ -values are significant, and the overall SPD shape and regions of high density are remarkably similar to the full 100% dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Clearly the method has performed conservatively in supporting the latter interpretation.

## 5. Population reconstruction results

The population reconstructions and associated global significance values for the 12 regions in this study, using the improved method, are shown in Fig. 3. We used the same null model as in the

previous study – an exponential fitted to the SPD from all 13,658 dates in the study area across the wider range of 10,000–4000 BP. As in our previous study the great majority of the regions show evidence of departures from the overall European exponential trend, with indications of population booms and busts, and eight of the 12 regions show strong indications of a statistically significant population increase immediately following the arrival of farming, with a ninth region (Little Poland) also showing some indication of



**Fig. 3.** Regional summed probability distributions (SPDs) of radiocarbon date analysis. Black line shows the SPD; dashed line shows a rolling 200 year mean of the SPD; dotted line shows the fitted null exponential model; red sections are local regions of unusually high density, blue shows regions of unusually low density; blue arrows represent approximate first farming dates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

this increase (the remaining three regions are Lowlands, Swedish Baltic Islands and Bohemia).

However, it is important to note that whilst many red regions appear to begin and end with periods of rapid growth and decline, the declines in most cases are not highlighted in blue to show significantly low density. This is unlike the results in our previous study (Shennan et al. 2013) where periods of significantly high density were in many cases soon followed by periods of significantly low density. This difference is almost certainly driven by the smaller sample sizes (mean = 313, range = 93–862, compared to the previous study mean = 623, range = 281–1732), and further explained in the demonstration section, in terms of evidence of absence. In any case strictly speaking, the *p*-value globally tests for a statistically significant departure from the null exponential model, whereas the interpretation of booms and busts are reasonable but subjective inferences drawn from the shape of the SPD and its highlighted time periods of ‘positive’ (red and blue) points. Therefore given the much smaller sample sizes in this study it seems more appropriate to interpret ‘bust’ as a marked fall in density, rather than a significantly low density in itself. As noted above, in all regions where the comparisons have been made, the population reconstructions have been supported by the lower resolution pollen evidence of changing anthropogenic impacts.

We can now turn to the individual regions. England and Wales without Wessex shows the same pattern as the other regions of the British Isles with a boom following the arrival of farming at c.6000 BP followed by a crash down to a level little more than half the preceding peak. The population remains at this relatively low level for nearly 800 years, starting to climb again at c.4500 BP, in a pattern closer to that of Ireland than Scotland or Wessex (Shennan et al. 2013, Fig. 3). The pattern for western France again indicates a population boom with the appearance of farming in the early 7th millennium BP. Here the peak is at around 6000 BP and it is apparent that the subsequent decline corresponds in time to the population expansion in England and Wales, possibly suggesting a link between the two regions (see discussion in Collard et al. 2010). Over the course of the next 800 years, apart from a brief uptick just after 5500 BP, the population proxy gradually drops to around two-thirds of its earlier peak at 5200 BP, a low point also observed in several other regions, though in France the rate of decline is less marked. The final drop at the end of the sequence may perhaps be related to a tendency to rely more on typology than radiocarbon dates among scholars working on the Bronze Age.

The Lowlands, Netherlands and Belgium, excluding the coastal regions subject to Holocene inundation, show a complex pattern of booms and busts, starting in the Mesolithic, with a peak at the start of the sequence at 8000 BP followed by a drop to half the peak level. A boom with the arrival of LBK farming groups in parts of the region at c.7300 BP is followed by another sharp drop in the early 7th millennium BP to less than half the LBK peak level; this corresponds to other evidence for the abandonment of the LBK areas of the Netherlands at this time (Bakels, 2007). Population then increases again as agriculture gradually infiltrates this region (Cappers and Raemaekers, 2008) before another drop in the second half of the 6th millennium BP, roughly synchronous with a drop in a number of other regions, as noted above. This in turn is followed by a rapid rise to peak at c.4700 BP, a level maintained for c.300 years, before a further sharp drop, though it cannot be excluded that different dating practices are relevant here.

Of the four Swedish regions, three indicate a population boom following the appearance of agriculture at around 6000 BP, followed by a rapid decline. However, while Western and Eastern Middle Sweden rise rapidly to a peak in the middle of the 6th millennium BP, in Central Southern Sweden the rate of growth is much slower, reaching a peak just after 5000 BP before declining to

half the peak level. Western Sweden shows a second peak at the same date, before declining very rapidly to less than one-third of the maximum. Eastern Middle Sweden declines steadily after the mid-6th millennium BP high point, to roughly half this level by 5000 BP. There is no evidence of a significant departure from expectation under the null model in the Swedish Baltic Islands. Three of the four Swedish regional patterns are very similar to those shown in Hinz et al. (2012); the exception is the Swedish Baltic Islands, for which the similarity is less clear, and which in any case is lacking in significance.

The two Polish regions, Kujavia and Little Poland, again show significant departures from the null. In Little Poland there is a rapid rise in population, reaching a peak just after 5000 BP, which is long after the initial appearance of LBK farmers in the region around 7500 BP. There is a further peak in the middle of the 5th millennium BP after a slight dip, and then a major fall-off to less than half the maximum, though it is possible that Bronze Age scholars take less radiocarbon samples. Kujavia has a more complex pattern, with a significant rise from below trend associated with the first farming, but the highest values occur in the mid 6th millennium BP, and are double the LBK levels. There is then a drop in the late 6th millennium BP, as in many other regions, before a short-lived peak at c.4800 BP.

In eastern Switzerland a major population boom is indicated with the arrival of farming in the late 7th millennium BP and a rapid decline in the mid 6th millennium BP, corresponding to that seen in the other three regions considered above. The subsequent rise to an equally high but more short-lived peak at c.4800 BP and the following major fall are corroborated by corresponding patterns in anthropogenic impacts inferred from pollen analysis for part of the same region (Lechterbeck et al. in press).

Of the two remaining regions, Bohemia shows no evidence of departure from the exponential trend. On the other hand, Moravia, which includes the adjacent area of Lower Austria, shows a dramatic population increase associated with the arrival of LBK farming c.7400 BP, followed by a rapid fall to little more than half the peak value just after 7000 BP, and then another immediate expansion to a peak at c.6600 BP before a collapse to a small fraction of the peak in the last centuries of the 7th millennium BP. We do not see a return to these previous high values, and there is a significant drop below trend in the early 5th millennium BP.

## 6. Conclusion

The density of radiocarbon dates in a dataset, and how this varies through time, provides a useful proxy for fluctuations in human population levels. This approach made a major step forward with the use of the new computational method presented in Shennan et al. (2013). For the first time statistical rigour was introduced using the MCSPD-method, providing confidence in whether the observed fluctuations could be considered significant, or were merely the consequence of sampling error and features in the calibration curve. This new tool allowed us to interrogate the EURO-EVOL database containing nearly 14,000 dates in unprecedented spatial detail, revealing boom and bust patterns that followed the first appearance of farming in many different regions of Europe.

This paper builds on that earlier groundwork, and introduces an improvement to the computational method, providing greater confidence that the patterns detected represent a genuine signal of changing population levels. The efficacy of this improved method is demonstrated by its ability to detect a statistically significant signal in remarkably small datasets—only 12 samples across 4000 years in the particular dataset tested. Equipped with this improved tool, we have therefore been able to assess 12 new and independent regions of the European Neolithic. Despite these new regions containing substantially fewer samples, we have again found evidence of

statistically significant fluctuations in regional population levels, occurring at different times in different regions. Our results provide compelling support for the argument proposed in the previous study, that boom-bust fluctuations rapidly followed the appearance of farming. The prevalence of this phenomenon in so many regions across Europe, combined with the dissimilarity and lack of synchronicity in the general shapes of the SPDs, supports the hypothesis of an endogenous, not climatic cause (cf. Müller, 2013).

### Acknowledgements

We are grateful to Robert Kelly for inviting us to speak in the 78th SAAs at The Frison Institute symposium “Dates as data: New applications of radiocarbon dating to archaeological problems” on 7th April 2013. We also thank Arek Marciniak, Peter Stadler and Magdalena Turkova for providing dates from Poland, Austria and Czech Republic respectively, and to the European Research Council for Advanced Grant #249390, ‘Cultural Evolution of Neolithic Europe’, that has supported this work.

### Appendix 1

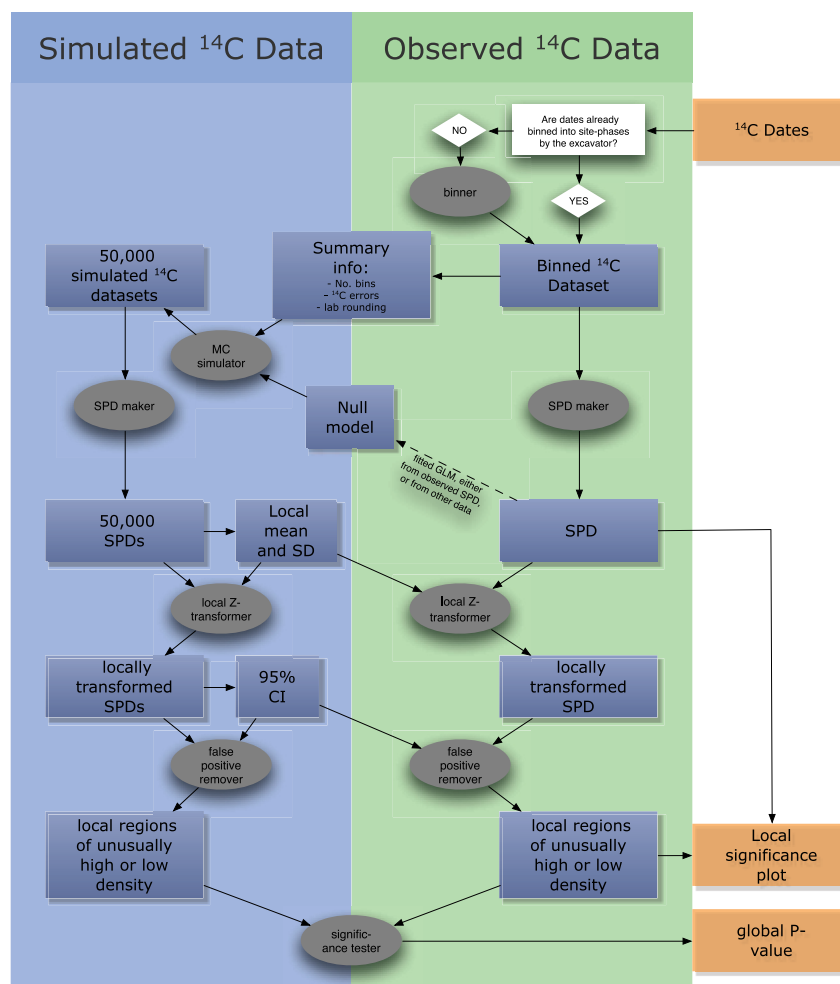
#### Details of Computational Method.

#### Binning the observed data:

If the excavator has not already grouped the  $^{14}\text{C}$  dates into specific site-phases, a systematic binning process is applied to dates from the same site using the ‘binner’ function. Once ordered, dates are only assigned to a new bin if there is at least a 200-year gap since the previous date. This addresses the excavator’s ‘ascertainment’ or ‘wealth’ bias, and ensures that each site-phase is equally weighted when generating the SPD. Tests suggest that the overall shape of the SPD is insensitive to the value used for the bin-width.

#### Generating a summed probability distribution (SPD) from $^{14}\text{C}$ data:

The ‘SPD maker’ function is used identically with both the observed and simulated data. Firstly, each  $^{14}\text{C}$  date and its associated error (the mean and standard deviation of the estimate) are calibrated in the usual manner using the intcal09 calibration curve (Reimer et al. 2009). Secondly, the resulting calibrated distributions from the same bin (or site-phase) are averaged. Finally the SPD from all bins are summed, trimmed to the time period of interest (8000–4000 BP), and normalised to unity. In the case of generating an SPD from simulated  $^{14}\text{C}$  data, only a single date is provided for



**Fig. 4.** showing a flowchart of the computational method. Rectangular boxes represent data at various stages of processing. Grey ellipses represent computational functions, with arrows showing data inputs/outputs to/from each function. Orange boxes represent initial data input and final output. Conceptually, identical computation is performed on both the observed data (green area) and the simulated data (blue area). Boxes on the boundary between blue and green represent data with properties of both. Ultimately it is the comparison between simulated and observed data that provides the basis of this method. Note: Not all calculations are represented, other simple calculations are used where data in one rectangle are transformed into data in another rectangle connected directly with an arrow (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

each bin (see ‘generating simulated  $^{14}\text{C}$  dates’), therefore the first averaging step is irrelevant. Tests show that when calibrating a single  $^{14}\text{C}$  date, the SPD maker function produces identically shaped distributions to those generated by other calibration software, such as OxCal (Bronk Ramsey, 2009).

#### Generating simulated $^{14}\text{C}$ dates:

The ‘Monte-Carlo (MC) Simulator’ function is at the heart of the computational method, and generates huge numbers of simulated  $^{14}\text{C}$  datasets. Each dataset is generated under the null model being tested, whilst being as similar as possible to the observed  $^{14}\text{C}$  dataset. In the pursuit of mimicking the observed data as closely as possible, this function is complex, and calls many helper functions. Firstly calendar dates are randomly generated under the null model, specifically,  $N$  dates are randomly sampled from the date range of interest, such that the probability of picking a particular date is determined by the null model; see ‘generating a null model’, where  $N$  is the number of bins in the dataset (In fact slightly more than  $N$  dates are required to sample from a slightly wider range, to allow for edge effects during the calibration process. These extra edges are subsequently trimmed by the ‘SPD maker’ function). Secondly each calendar date is converted into a single  $^{14}\text{C}$  date by ‘back-calibrating’ through the intcal 2009 curve; specifically, a random  $^{14}\text{C}$  date is picked from the intcal curve error at the corresponding calendar date. Thirdly a simple Monte-Carlo method is used to apply a rounding error to each simulated  $^{14}\text{C}$  date. This mimics the rounding applied by convention (Stuiver and Pollach, 1977) to  $^{14}\text{C}$  dates by many radiocarbon laboratories. Specifically, the observed dataset is probed to estimate the proportion of dates that were rounded to the nearest 1 yr, 5 yrs and 10 yrs. This is used to round the same proportion of simulated dates to 1, 5 or 10 years. Finally the simulated  $^{14}\text{C}$  errors (standard deviations) are generated by sampling with replacement  $N$  errors from the observed  $^{14}\text{C}$  errors.

#### Generating a null model:

The choice of exactly which null model to be tested is dependent on the specific research question being asked. Nevertheless, given the background expectation of human population increase, and the taphonomic losses of material through time, an exponential model is often appropriate. In some circumstances other shapes may be more appropriate, for example a sigmoid null for data from an island where it appears carrying capacity has been reached, and population levels are known to have been constant for a long time. Once a general shape is selected, a conservative approach to estimating the exact shape parameters of this null requires fitting it to the observed SPD as closely as possible. This model fitting uses standard methods of least square regression and a generalised linear model (GLM). Alternatively, as was the case in this paper and in Shennan et al. (2013), an exponential null model was fitted to the combined dataset of all regions in the database, over the longer period of 10,000–4000 BP, again using a GLM. The rationale for this second approach comes from the advantages brought by the law of large numbers; the larger the sample size, the closer the distribution will approximate the true distribution of the overall European population level through time. Once the parameters (and therefore the exact shape) of the null have been estimated, the total area is the normalised to unity, so that it can be directly used as a vector of probabilities by the ‘MC simulator’ function.

#### Transforming the calibrated SPDs:

In order to assess local differences in the density of a SPD it is first necessary to de-trend the SPD both from its (exponential) null

model, and from spurious short term wiggles (less than 200 years) generated by the calibration process. This is achieved using a local Z-score transformation. Specifically, a vector of means and a vector of SDs are calculated from all the simulated SPDs. These two vectors are then used by the ‘local Z-transformer’ to de-trend both the observed SPD and each simulated SPD. Local 95% confidence intervals are then calculated from the transformed simulated SPDs.

#### Finding local periods of unusually high or low density

The 95% confidence interval represents the range in which we can expect 95% of the data to occupy. Local sections of the observed SPD that fall outside the confidence interval can be considered unusually high or low density, and plotted as such as red or blue. However, by definition, approximately 5% of any SPD can be expected to fall outside this confidence interval from pure chance, which can be considered ‘false positive’ sections. In the case of a simulated SPD these false positives cannot be considered to represent any interesting or significant feature, since by definition the simulated SPD was generated through random processes. Similarly, approximately 5% of the observed SPD can also be expected to fall outside the confidence interval. Therefore we apply the ‘false positive remover’ function to filter out most likely candidates of false positives. This function uses the principle that since false positives are random events, we can expect them to be independently distributed through a SPD, therefore more likely to occur alone than in pairs, more likely in pairs than triplets etc., whereas ‘true positives’ are more likely to have low entropy, exhibit order, and therefore occur in blocks. This function therefore applies repeated cleaning sweeps to the SPD, removing single occurrences of a positive followed by occurrences of a positive with only one positive neighbour, until a maximum of 5% of the SPD has been demoted from the category of unusually high/low density. This is performed identically to the observed SPD, and to each simulated SPD.

#### Calculating global significance

Using the ‘significance tester’ function, a summary statistic (the total area of the remaining positives that are outside the confidence interval) is calculated both for the observed transformed SPD and for each simulated transformed SPD. A global  $P$ -value is then simply calculated as the proportion of simulations for which their summary statistic is as (or more) extreme than the observed summary statistic.

## Appendix 2

### Radiocarbon data sources

Gamble, C., Davies, W., Pettitt, P., Hazelwood, L. & Richards, M. The archaeological and genetic foundations of the European population during the Late Glacial: implications for ‘agricultural thinking’. *Camb. Archaeol. J.* 15, 193–223 (2005).

Guilaine, J. 1998. Atlas du néolithique européen. Volume 2A & 2B, L’Europe occidentale, Université de Liège, Liège.

Weninger, B. 2009. In Crombé, P., Van Strydonk, M., Boudin, M., Batz, M (Eds.), Chronology and Evolution in the Mesolithic of Northwest Europe, Cambridge Scholars Publishing, Cambridge, pp. 143–176.

Steele J., Shennan, S.J. 2003 Spatial and Chronological Patterns in the Neolithisation. [Internet] [http://archaeologydataservice.ac.uk/archives/view/c14\\_meso/](http://archaeologydataservice.ac.uk/archives/view/c14_meso/).

Galate, P. 2011. BANADORA, Banque de données des dates radiocarbones de Lyon pour l’Europe et le Proche-Orient. Université Claude Bernard, Lyon. [Internet] <http://www.archeometrie.mom.fr/banadora/>.



Council for British Archaeology. 2008. Archaeological Site Index to Radiocarbon Dates from Great Britain and Ireland. [Internet] [http://archaeologydataservice.ac.uk/archives/view/c14\\_cba/](http://archaeologydataservice.ac.uk/archives/view/c14_cba/).

Hinz, M. 2012. RADON – Radiocarbon dates online 2012. Central European database of <sup>14</sup>C dates for the Neolithic and Early Bronze Age, University of Kiel. Available at: <http://radon.ufg.uni-kiel.de>.

Vermeersch, P. 2011. Radiocarbon Palaeolithic Europe Database v13. Dept. of Earth and Environmental Sciences, Katholieke Universiteit Leuven. [Internet] <http://www.ees.kuleuven.be/geography/projects/14c-palaeolithic/index.html>.

## Appendix 3

Approximate first farming dates.

Region	BCE	Reference
England and Wales without Wessex	4000	Whittle et al. (2011)
Western France	4800	Marchand (2007)
Lowlands	5200	Amkreutz et al. (2012)
Western Sweden	4000	Sørensen and Karg (2014)
Central Southern Sweden	4000	Sørensen and Karg (2014)
Eastern Middle Sweden	4000	Knutsson and Knutsson (2003)
Swedish Baltic Islands	4000	Eriksson et al. (2008)
Kujavia	5400	Pyzel (2009)
Little Poland	5400	Pyzel (2010)
Eastern Switzerland	4400	Denaire et al. (2011)
Bohemia	5400	Compromise of 5500 and 5300, from <a href="http://www.bylany.com/eng/bylany_chronology.html">http://www.bylany.com/eng/bylany_chronology.html</a> and Stadler and Kotova (2010)
Moravia	5500	Oross and Banffy (2009)

Appendix 3 provides the sources for the approximated first farming dates shown in Fig. 3 using blue arrows. Conversion from BCE to BP was achieved by adding 1950 years.

## References

- Amkreutz, L., Bakels, C., Brounen, F., Dijkman, W., Van Gijn, A., De Grooth, M., Hendrix, W., De Reus, T., Schmitz, H., Van De Velde, P., Verbaas, A., Vromen, H., De Warrimont, J.P., Van Wijk, I., 2012. An Odyssey along the river Meuse: new perspectives on old Dutch LBK research. *Notae Praehist.* 32, 51–71.
- Ammerman, A.J., Cavalli-Sforza, L.L., 1971. Measuring the rate of spread of early farming in Europe. *Man* 6, 674–688.
- Bakels, C., 2007. Nature or culture? Cereal crops raised by Neolithic farmers on Dutch loess soils. In: Colledge, S., Conolly, J. (Eds.), *The Origins, Spread and Use of Domestic Plants in Neolithic Southwest Asia and Europe*. Left Coast Press, Walnut Creek.
- Bamforth, D.B., Grund, B., 2012. Radiocarbon calibration curves, summed probability distributions, and early Paleoindian population trends in North America. *J. Archaeol. Sci.* 39, 1768–1774.
- Bocquet-Appel, J.P., 2002. Paleoanthropological traces of a Neolithic demographic transition. *Curr. Anthropol.* 43, 637–650.
- Bronk Ramsey, C., 2009. Bayesian analysis of radiocarbon dates. *Radiocarbon* 51, 337–360.
- Cappers, R., Raemaekers, D., 2008. Cereal cultivation at Swifterbant? Neolithic wetland farming on the North European plain. *Curr. Anthropol.* 49, 385–402.
- 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. <http://dx.doi.org/10.1016/j.jas.2014.05.030>.

- Denaire, A., Doppler, T., Nicod, P.Y., Van Willigen, S., 2011. Espaces Culturels, Frontières et Interactions au 5ème Millénaire entre la Plaine du Rhin Supérieur et les Rivages de la Méditerranée. *Annu. Archéol. Suisse* 94, 21–59.
- Eriksson, G., Linderholm, A., Fornander, E., Kanstrup, M., Schoultz, P., Olofsson, H., Lidén, K., 2008. Same island, different diet: cultural evolution of food practice on Öland, Sweden, from the Mesolithic to the Roman Period. *J. Anthropol. Archaeol.* 27, 520–543.
- Hinz, M., Feeser, I., Sjögren, K.G., Müller, J., 2012. Demography and the intensity of cultural activities: an evaluation of funnel beaker societies (4200–2800 cal BC). *J. Archaeol. Sci.* 39, 3331–3340.
- Knutsson, H., Knutsson, K., 2003. Stone age transitions. Neolithisation in Central Scandinavia. *Doc. Praehist.* XXX, 48–78.
- Lechterbeck, J., Edinborough, K., Kerig, T., Fyfe, R., Roberts, N., Shennan, S., 2014. Is Neolithic land use correlated with demography? An evaluation of pollen derived land cover and radiocarbon inferred demographic change from Central Europe. *Holocene* (in press).
- Li, H., Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475, 493–497.
- Marchand, G., 2007. Sur les rives de l'Atlantique: mégalithes et enceintes 4800–3500. In: Demoule, J. (Ed.), *La révolution néolithique en France. La Découverte*, Paris, pp. 97–120.
- McEvedy, C., Jones, R., 1978. *Atlas of World Population History*. Penguin, London.
- Müller, J., 2013. Demographic traces of technological innovation, social change and mobility: from 1 to 8 million Europeans (6000–2000 BCE). In: Kadrow, S., Włodarczak, P. (Eds.), *Environment and Subsistence – Forty Years after Janusz Kruk's Settlement Studies*. Institute of Archaeology UR & Verlag Dr. Rudolf Habelt GmbH, Rzeszów, pp. 493–506.
- Oross, K., Banffy, E., 2009. Three successive waves of Neolithisation. LBK development in Transdanubia. *Doc. Praehist.* XXXVI, 175–189.
- Pyzel, J., 2009. Settlement history of the Linear Band Pottery culture in Kuyavia. In: Hofmann, D., Bickle, P. (Eds.), *Creating Communities: New Advances in Central European Neolithic Research*. Oxbow, Oxford, pp. 71–79.
- Pyzel, J., 2010. Zofipole/ackovy/Flomborn. On the problems of polish subphase Ib of the linear Band Pottery culture. In: Suteková, J., Pavúk, P., Kalábková, P., Kovár, B. (Eds.), *PANTA RHEI. Studies in the Chronology and Cultural Development of South-eastern and Central Europe in Earlier Prehistory Presented to Juraj Pavúk on the Occasion of his 75. Birthday*, pp. 539–547.
- Reimer, P.J., Baillie, M.G.L., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Bronk Ramsey, C., Buck, C.E., Burr, G.S., Edwards, R.L., Friedrich, M., Grootes, P.M., Guilderson, T.P., Hajdas, I., Heaton, T.J., Hogg, A.G., Hughen, K.A., Kaiser, K.F., Kromer, B., McCormac, F.G., Manning, S.W., Reimer, R.W., Richards, D.A., Southon, J.R., Talamo, S., Turney, C.S.M., van der Plicht, J., Weyhenmeyer, C.E., 2009. *IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP*. *Radiocarbon* 51 (4), 1111–1150.
- Rick, J.W., 1987. Dates as data: an examination of the Peruvian Preceramic Radiocarbon Record. *Am. Antiq.* 52, 55–73.
- 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, 2486.
- Sørensen, L., Karg, S., 2014. The expansion of agrarian societies towards the north – new evidence for agriculture during the Mesolithic/Neolithic transition in Southern Scandinavia. *J. Archaeol. Sci.* 51, 98–114. <http://dx.doi.org/10.1016/j.jas.2012.08.042>.
- Stadler, P., Kotova, N., 2010. The early Neolithic settlement from Brunn Wolfholz in lower Austria and the problem of typology and chronology of the low bowls with sharply curved wall. *Slov. Acheol.* LVIII, 207–228.
- Stuiver, M., Pollach, H., 1977. Discussions of reporting <sup>14</sup>C data. *Radiocarbon* 19, 355–363.
- Surovell, T.A., Byrd Finley, J., Smith, G.M., Brantingham, P.J., Kelly, R., 2009. Correcting temporal frequency distributions for taphonomic bias. *J. Archaeol. Sci.* 36, 1715–1724.
- Whitehouse, N.J., Schulting, R.J., McClatchie, M., Barratt, P., McLaughlin, T.R., Bogaard, A., Colledge, S., Marchant, R., Gaffrey, J., Bunting, M.J., 2014. Neolithic agriculture on the European western frontier: the boom and bust of early farming in Ireland. *J. Archaeol. Sci.* 51, 181–205. <http://dx.doi.org/10.1016/j.jas.2013.08.009>.
- Williams, A., 2012. The use of summed radiocarbon probability distributions in archaeology: a review of methods. *J. Archaeol. Sci.* 39, 578–589.
- Woodbridge, J., Fyfe, R.M., Roberts, N., Downey, S., Edinborough, K., Shennan, S., 2014. The impact of the Neolithic agricultural transition in Britain: a comparison of pollen-based land-cover and archaeological <sup>14</sup>C date-inferred population change. *J. Archaeol. Sci.* 51, 216–224. <http://dx.doi.org/10.1016/j.jas.2012.10.025>.
- Whittle, A., Healy, F., Bayliss, A., 2011. *Gathering Time: Dating the Early Neolithic Enclosures of Southern Britain and Ireland*. Oxbow, Oxford.
- Zimmermann, A., Richter, J., Frank, T., Wendt, K.-P., 2004. *Landschaftsarchäologie II – Überlegungen zu Prinzipien einer Landschaftsarchäologie*. *BerGK* 85, 37–95.