Parsing prehistoric patterns: Prospects and limitations of a big radiocarbon dataset for understanding the impact of climate on Late Palaeolithic and Mesolithic populations in northwest Europe (16–7.5 ka calBP)

P.W. Hoebe *, J.H.M. Peeters, S. Arnoldussen

Groningen Institute of Archaeology, University of Groningen, Poststraat 6, 9712ER, Groningen, the Netherlands

ARTICLE INFO

Keywords: Mesolithic Late Palaeolithic Climate change Radiocarbon data Bias exploration Northwest Europe

ABSTRACT

Hunter-gatherer populations in northwest Europe were variably affected by Late Glacial and Early Holocene climate fluctuations and their effects on sea level and the environment. We investigate the impact of these fluctuations with a dates-as-data approach to a large radiocarbon dataset. Radiocarbon dates are used as a proxy for past human activity, the intensity, nature and archaeological visibility of which will indirectly influence date density. The significance of changes is explored using Kernel Density Estimates and model tested Summed Probability Distributions. Whereas previous studies have focused on smaller highly curated datasets to minimise research and preservation biases, our more inclusive approach maximises sample size, which is essential for these methods to reliably reflect underlying patterns. To deal with biases, we test subsets of the dataset that are potentially affected by differences in formation processes.

The summed radiocarbon dataset follows the general fluctuations of climate conditions, showing increased activity in temperate periods and decreased activity during cold phases. Our results indicate significant periods of interest where the data deviates positively or negatively from our models. Notably we observe the impact of the Younger Dryas, Preboreal Oscillation and the 8.2 ka event on the density of hunter-gatherer activity. Additionally we see peaks in activity in our dataset during the Early and Late Boreal. Permutation testing of different regions in the research area shows these patterns are geographically differentiated.

Our exploration of biasing factors indicates that we should be careful to interpret the abovementioned patterns, as different sampling processes and national policies may lie at the basis of several patterns. Furthermore, calibration artefacts may also cause issues at key parts of the timeline. Dates-as-data approaches require an understanding of the archaeology, the timing of external events, the impact of the calibration curve and how biases inherent to the dataset and research area may have influenced the formation of patterns in the result.

1. Introduction

Hunter-gatherer populations in northwest Europe were variably impacted by Late Glacial and Early Holocene climate fluctuations (Bond et al., 1997, 2001; Clark et al., 2009; Rasmussen et al., 2014; Patton et al., 2017), as well as their effects on sea level and the environment (Weninger et al., 2008; Patton et al., 2017; Blankholm, 2020). Rapidly changing conditions presented both challenges and opportunities for human societies, resulting in the enactment of region-specific strategies. A ‘big data’ approach to the archaeological record is used here as an initial exploration of the timing of changes in human activity (Gattiglia, 2015; Huggett, 2020). Specifically we analyse fluctuations in the summed probability distributions of radiocarbon dates as a proxy for changes in the intensity and nature of human activity (Crema and Bevan, 2020), and juxtapose this with the timing of changing external conditions. It has been demonstrated elsewhere, and for younger periods, that such approaches can provide valuable insights into large-scale patterns of human behaviour such as climate-related adaptation, or societal collapse (Palmisano et al., 2017; Bevan et al., 2018; Lawrence et al., 2021). A large spatiotemporal framework is required to analyse human-environmental interaction and aspects like resilience, adaptation or collapse in general, and specifically for nomadic and traditional prehistoric hunter-gatherer communities.

For the northwest European Late Palaeolithic and Mesolithic, studies

* Corresponding author.
E-mail address: p.w.hoebe@rug.nl (P.W. Hoebe).

https://doi.org/10.1016/j.jasrep.2023.103944
Received 2 October 2022; Received in revised form 16 February 2023; Accepted 5 March 2023
2352-409X/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
have mostly focused on demographic effects of specific climate events, or concentrated on a limited research area or period (Riede, 2009; Vermeersch, 2011; Robinson et al., 2013; Wicks and Mithen, 2014; Crombé and Robinson, 2017; Waddington and Wicks, 2017; Griffiths and Robinson, 2018; Van Maldegem et al., 2021). These studies involved smaller, curated datasets, resulting in relatively small sample sizes. While focusing on smaller regions may serve to maintain taphonomic and ecological coherence (Ward and Larcombe, 2021), sample sizes of such case studies may not be high enough to reliably reflect events (Hinz, 2020). We therefore adopt a sufficiently large spatiotemporal framework to minimise sampling error, though it should be noted that such an approach results in a “space-averaged” estimate that might not be representative of its subregions (Crema, 2022).

Besides investigating the impact of climate on hunter-gatherer activity, this paper aims at a critical assessment of the prospects and limitations of analysing and interpreting a ’big’ radiocarbon dataset, consisting of 5415 dates across 2012 sites. As part of the first author’s PhD project on climate and environment-driven sociocultural change among Late Palaeolithic and Mesolithic hunter-gatherers (16-7.5 ka calBP) in the region surrounding Doggerland (i.e. the southern North Sea), this dataset is used to investigate the nature of spatiotemporal patterning in terms of behavioural dynamics. We test hypotheses about the overall growth of the archaeological record, using Kernel Density Estimates (KDE) and model tested Summed Probability Distributions (SPD) (Bronk Ramsey, 2017; Brown, 2017; Crema and Bevan, 2020). We assess the interpretive strength of the results through permutation tested SPDs of subsets of the data in order to explore issues of preservation- and research bias inherent to big datasets (Williams, 2012; Contreras and Meadows, 2014; Becerra-Valdivia et al., 2020; Ward and Larcombe, 2021). This exploration is a critical step in defining the possibilities and limitations for the subsequent stages of research in our goal of understanding the impact of climate and environment-driven sociocultural change among Late Palaeolithic and Mesolithic hunter-gatherers in the region surrounding Doggerland.

2. Human activity, environment and climate

People live in diverse environments across the globe, due to our ability to adapt to (extreme) circumstances through cultural adaptations including subsistence and social organisation. Early prehistoric communities must have similarly adapted to the sometimes extreme circumstances of past climate and environmental changes, shaping or constraining their behaviour (Kelly et al., 2013; Barton et al., 2018; Degroot et al., 2022; Ordonez and Riede, 2022). Changes in key climate variables such as temperature and precipitation are at the basis of the spatial distribution of climate zones and related biomes or habitat types (i.e. tundra, boreal forest, and temperate forest, Stuart Chapin et al., 2011: 50-59). Across these different biomes there are significant differences in the net productivity of primary, and in turn secondary biomass (i.e. plant and animals respectively), as well as overall species diversity (Binford, 2001: 79-109; Stuart Chapin et al., 2011:59-60; Ordonez and Riede, 2022). Different environmental conditions across biomes require different cultural adaptations from the human communities inhabiting them. As climate repeatedly changed drastically and rapidly during the Late Glacial and Early Holocene (Bond et al., 1997, 2001; Nesje et al., 2004; Clark et al., 2009; Rasmussen et al., 2014; Patton et al., 2017), we expect this caused range shifts in both plant and animal communities.

We hypothesise (Table 1) how climate regime shifts between stadials and interstadials may have led to the movement of people to more desirable areas, decreased (famine, disease, dispersal) or increased population density (convergence), or changing sociocultural and demographic strategies (birth-rate change, population control, territorial conflict, food storage and change in subsistence strategies). Changes in the geographical occurrence of various types of activities, as well as the nature of activities, are expected to be reflected in the archaeological record.

In addition we hypothesise that cold and/or dry phases within interstadials similarly affected the distribution or the productivity of biomes, in turn affecting human activity. The phases considered are GI-1d, GI-1c2, GI-1b during the BA-Interstadial (Rasmussen et al., 2014) and the Preboreal Oscillation / 11.4 ka event (Björck et al., 1997), 10.3 ka event (Bond et al., 1997, 2001), 9.3 ka event (Yu et al., 2010), and the 8.2 ka event (Alley and Agüestdöttir, 2005) during the Early Holocene. Declined human activity has already been established for the 8.2 ka event/Storegga Slide Tsunami in Northern case studies (Manninen, 2014; Wicks and Mithen, 2014; Waddington and Wicks, 2017; Jørgensen, 2020; Mithen and Wicks, 2021).

Finally, rising sea level culminating in the flooding of Doggerland (Coles, 1998; Ward et al., 2006; Weninger et al., 2008; Cohen et al., 2022).

### Table 1

<table>
<thead>
<tr>
<th>Climate Phase</th>
<th>Region</th>
<th>Conditions</th>
<th>Habitat Type</th>
<th>Human Activity</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>O. Dryas GS-2</td>
<td>Stadial</td>
<td>– – –</td>
<td>Subpolar desert, steppe tundra</td>
<td>Restricted to southern regions, followed by northward expansion</td>
<td>1,2</td>
</tr>
<tr>
<td>Balling GI-1e</td>
<td>Transition</td>
<td>++ + +</td>
<td>Tundra</td>
<td>Dispersed northward expansion</td>
<td>1–4</td>
</tr>
<tr>
<td>Allerød GI-1c - GI-1a</td>
<td>Interstadial</td>
<td>– ++ ++</td>
<td>Birch–pine forests</td>
<td>Dense northward expansion, population increase</td>
<td>2, 5–7</td>
</tr>
<tr>
<td>E. Y. Dryas - Preboreal</td>
<td>Transition</td>
<td>– – –</td>
<td>Birch–pine forests</td>
<td>Persistence</td>
<td>2, 5–10, 14, 16</td>
</tr>
<tr>
<td>Boreal - Atlantic</td>
<td>Interstadial</td>
<td>– ++ ++</td>
<td>Temperate forests</td>
<td>Population expansion and increase, cultural change</td>
<td>12–16</td>
</tr>
</tbody>
</table>

---

3. Limitations of dates-as-data approaches

Dates-as-data approaches have a long research history (Williams, 2012; Carleton and Groucutt, 2021) in both palaeoenvironmental and archaeological studies, hailing back to the late 1960s. For archaeology, Rick’s influential ‘Dates as data’ paper (Rick, 1987) popularised the approach, highlighting its potential as a proxy for studying past population dynamics, but also calling attention to biasing factors. Refinement of the method has led to their common application in the last decade with the increased availability of large radiocarbon databases and calibration programs (Williams, 2012). Current research focuses on Summed Probability distributions (SPDs) or Kernel Density Estimates (KDEs) of large numbers of dates (Bronk Ramsey, 2017; Brown, 2017; Crema and Bevan, 2020; Crema, 2022).

However, limitations and issues in dates-as-data approaches persist. Inherent to archaeology as a whole is the nature of our data as a non-random sample of past human activity (Fig. 1). The formation of the potential archaeological radiocarbon record itself is influenced by the nature and intensity of human activity, preservation conditions and taphonomic processes (Ward and Larcombe, 2021). Which portion of the record is excavated, researched and eventually submitted for radiocarbon dating depends on non-random factors such as accessibility due to regional differences in sedimentation history (ibid), national differences in research policy and focus (Hinz et al., 2012; Torfing, 2015; Crema, 2022), making comparison across regions challenging. Some scholars focus on correcting for such biases (Surovell et al., 2009; Onnen et al., 2010; Timpson et al., 2014; Bluhm and Surovell, 2019; Crema and Bevan, 2020), while others advise against frequency analysis approaches to radiocarbon dates altogether (Contreras and Meadows, 2014; Beccera-Valdivia et al., 2020; Ward and Larcombe, 2021).

In addition to research and preservation biases, there are biases inherent to the method. Fluctuations in the calibration curve form an intrinsic concern for the statistical methods used to summarise large radiocarbon datasets (Williams, 2012; Bronk Ramsey, 2017; Weninger and Edinborough, 2020). How well SPDs reflect past demographic patterns has been explored through modelling studies (Contreras and Meadows, 2014; Bronk Ramsey, 2017; Hinz, 2020). These show that changes in SPD density can be difficult to interpret, due to calibration curve irregularities. Bronk Ramsey (Bronk Ramsey, 2017) advises using KDE as an alternative. This method may be more reliable than SPD as the shape of the distribution is not strongly influenced by the details of the calibration curve (Bronk Ramsey, 2017; Brown, 2017). Consequently, KDE plots are less noisy than SPDs, making them better suited for the identification of underlying signals (Bronk Ramsey, 2017). Another solution is to test SPDs against simulated envelopes of hypothesised growth models (Shennan et al., 2013; Timpson et al., 2014; Crema and Bevan, 2020). Irregularities are amplified if redundant normalisation is applied in the process, causing spikes in the SPD (Fig. 3c): these spikes are statistical artefacts that appear at steep sections of the calibration curve (Weninger et al., 2015; Crema and Bevan, 2020). They indicate changing rates of $^{14}C$ production in the atmosphere caused for example by cosmic rays or solar storms (Reimer et al., 2020; Heaton et al., 2021), both of which can affect or coincide with climate change. Simulated model test envelopes also display offset dips at these steep sections, which are amplified when dates are normalised (Crema and Bevan, 2020). Fluctuation in an SPD’s density itself is influenced by the same effects, and should therefore not be interpreted without reference to a modelled hypothesis (Bronk Ramsey, 2017; Crema and Bevan, 2020). To estimate the likelihood that an SPD will reflect underlying declines in activity of differing severity, sample size (Timpson et al., 2015; Bronk Ramsey, 2017) and data density must be sufficient (Hinz, 2020).

4. Dataset preparation and general frequency distributions

4.1. Collating and restructuring

The radiocarbon data was assembled by combining existing datasets from the United Kingdom’s ADS, CalPal, EuroEvol, PACEA, RADON and the Palaeolithic Europe radiocarbon database (see Supplementary Information document: section 1 and 2 - SI-1: Source overview, SI-2: Dataset). Because of data accessibility, as well as national differences in practices regarding open access, data density varies by country (SI-4). To improve the spatial and temporal representativity, additional data was collected from publications (Niekus, 2005; Grimm and Weber, 2008; Maier, 2015; Peeters and Niekus, 2017; Waddington and Wicks, 2017; Griffiths and Robinson, 2018; Gehlen et al., 2020; Jensen et al., 2020). Variations in quality and structure are important issues when integrating disparate radiocarbon datasets as these are not always supplied with re-use in mind. Consequently, a portion of the information had to be extracted, transformed, and redistributed into separate columns, most pertinently data that had to be quantified such as site names, material categories and species (SI-5).
4.2. Methods and visualisation of frequency distributions

After deselecting dates without reported lab-codes and error margins exceeding 200 years,\(^1\) calibration was done with the R package ‘rcarbon’ v1.4.2 (Crema and Bevan, 2021) using Intcal20 (Reimer et al., 2020). Dates exceeding the timeframe (16–7.5 ka calBP) were included in this to avoid edge effects, but these are excluded from the final counts (total = 6475, 1060 outside timeframe). 5415 calibrated dates from 2012 sites fall within the timeframe (SI-4). Besides sites, we distinguish site phases to get a better sense of the intensity of human activity through time. Phases were determined using the `binPrep()` function, which is also used to prepare a dataset for SPD aggregation. Binning counteracts overrepresentation caused by research focus, aggregating (for SPDs) or sampling (for KDEs) dates that are close in time within the same phase or site (Shennan et al., 2013; Crema and Bevan, 2020). While binning can skew results if dates are inaccurate (Becerra-Valdivia et al., 2020), manual assessment of data accuracy at site or phase level is unfeasible for large heterogeneous datasets. The effect of bin size was visualised using `binSense()` and set at 100 years (see Fig. 3b), resulting in 3802 site phases. To assess the temporal distribution of date and site-phase density, we plotted the aoristic weight per century for each climate chronozone (Fig. 2). This was calculated by assigning portions of the calibrated date ranges of each date or phase to each chronozone, then dividing by the duration of the chronozone (SI-6).

Both SPD and KDE were used to visualise the general frequency distribution. The `spd()` function aggregates calibrated radiocarbon dates and sums probabilities for each year, producing a plot that shows fluctuating density through time (Crema and Bevan, 2020; Crema, 2022). For the analysis SPDs were made without redundant normalisation (see section 3), but the effect of normalisation is visualised in Fig. 3c.

The KDE was made with rcarbon’s `sampledates()` and `ckde()` functions (Fig. 3d) with bins and bootstrapping enabled. Calibrated calendar dates were randomly sampled from each probability density function and a kernel density estimate with a user-defined bandwidth was generated. Bandwidths of several decades are preferred in archaeology, so sudden changes remain visible. We chose a bandwidth of 50 years, allowing the visualisation of changes at the generational scale (see SI-6 for results at different bandwidths). The KDE was generated over 500 simulations and visualised as an envelope (Brown, 2017; Rowan McLaughlin, 2019; Crema and Bevan, 2020). The SPD (Fig. 3c) and KDE (Fig. 3d) show higher densities during the BA Interstadial and the Holocene (Fig. 3a). Periods of decline are identifiable following and during key climate deteriorations (GS1, 11.4, 8.2). These patterns, and others that may not be directly related to climate are discussed in section 5.

As unreliable dates in large datasets may impact the overall pattern, we ran a vetting procedure to explore the impact on the temporal distribution (SI-6). Approaches to vetting differ across the field. While some focus on maximizing sample size and correcting for biases through computational methods (Surovell et al., 2009; Williams, 2012; Timpson et al., 2015), others argue for vetting the dataset based on certain standards prior to computational analysis (Torfing, 2015; Becerra-Valdivia et al., 2020; Van Maldegem et al., 2021; Ward and Larcombe, 2021). While the former has the disadvantage of retaining erroneous data that may skew results, the latter has the disadvantage of reducing sample size and introducing a new source of bias: the researcher’s judgement in the deselection process (Timpson et al., 2015).

The effects of the vetting procedure (SI-6) demonstrated that 1) the dataset was reduced drastically by 1869 dates and 762 sites across the temporal framework; 2) vetting impact was not uniform, affecting some periods more than others and causing density fluctuations of higher magnitude and 3) change in the timing of density fluctuations was minimal. While the latter is reassuring, the non-uniform impact makes the vetted dataset a less conservative option for model testing. Additionally, density reduction increases sampling error and makes it less likely for the method to reliably detect underlying patterns (Hinz, 2020). Based on this we proceed with the full dataset for further analysis.

5. Analysis of frequency distributions

5.1. Methods and hypotheses for model testing

SPDs should be tested against models that hypothesize different patterns of growth (Shennan et al., 2013; Crema et al., 2016; Bevan et al., 2018; Crema and Bevan, 2020; Palmisano et al., 2021; Crema, 2022). In rcarbon, the Monte-Carlo `modeltest()` function, compares the observed SPD against a distribution of expected SPDs given a growth model (Crema and Bevan, 2020; Crema, 2022). The function takes random samples from the model, uncalibrates, recalibrates and
aggregates the dates, resulting in an expected SPD. To account for variations arising from sampling error and the calibration curve, the process is repeated many times (500 simulations), generating a distribution of SPDs, visualised as an envelope which is compared to the observed SPD to identify significant time intervals that deviate from the null-hypothesis. A running mean of 50 years was used to decrease noise. Model tests were performed with an exponential model and a custom climate model based on interpolated NGRIP data (North Greenland Ice Core Project, 2004; Bazin et al., 2013).

In such model tests, the null-hypothesis is (almost) always rejected, as strict correlation of the observed data with any modelled growth pattern is unexpected. What matters in the interpretation of the results is when and how much the data deviates from the null-hypotheses (Crema, 2022). The exponential model test’s hypothesis is uninterrupted growth across the timeframe (Shennan et al., 2013). If human activity decreased during cold phases, this should be reflected as significant negative

**Fig. 2.** Above: spatial distribution of the dataset (n = 2012 sites); below: aoristic weighing (n dates per year per chronozone) of a. radiocarbon dates (n = 5415) and b. site phases (n = 3802) per climate chronozone in the temporal frame.
deviations from the model during or following such an event. Positive deviations would suggest periods of increased activity that may be of interest. The climate model test’s hypothesis is that human activity will fluctuate with temperature. Instead of building a more elaborate model that includes assumptions about the specific relationships between climate, environment, and human behaviour, we chose to keep this simple, using temperature proxy data (NGRIP $\delta^{18}O$) as direct input for the construction of a distribution of expected SPDs. This way, differences in the timing of patterns can be made more explicit, possibly identifying delayed responses.

First, the correlation of the temperature proxy and the SPD was demonstrated by the correlation coefficient ($r_s = 0.902, p < 0.000$) and the alignment of modes of the frequency distributions of both datasets after standardization (Fig. 4, SI-7). The NGRIP data was fitted to the scale of the SPD (SI-7 and ‘scripts/Modeltest climate.R’) so the timing of changes in temperature could be compared to the timing of changes in the SPD. The SPD’s mode during the Dryas stadials is assumed to be appropriate for the mode of contemporaneous temperature levels, as well as the modes during the BA-interstadial and Early Holocene. Basically, this test assumes relatively close adaptation to temperature conditions during different climate regimes. As shown in Fig. 4, SPD values approximately follow the same trend as the temperature proxy, especially during the stadial regimes and the (transition to the) Early Holocene. The test will then highlight periods in which this assumption is not met. Deviations might be caused by environmental changes that are not reflected in temperature proxies (e.g. sea level rise) or socio-cultural changes. Different modes of hunter-gatherer activity during the climate regimes of the Late Glacial and the Early Holocene have been established for southwest Europe, establishing adaptation through changes in the social ecological technical system (Barton et al., 2018).

Using NGRIP has disadvantages, as the timing and intensity of $\delta^{18}O$ fluctuations are influenced by multiple factors including temperature, moisture and ice sheet extent (North Greenland Ice Core Project, 2004). Besides, shifts in Greenland conditions do not necessarily mean similar changes in northwest Europe, let alone regional variability within the research area. However, a general proxy is better suited for our purpose than attempting to align multiple regional proxies such as pollen records, which would not only involve a complex study in itself, but also...
introduce a new range of biases and alignment issues. In terms of timing, depending on the environmental effects of any change, up to two centuries of delayed response between the NGRIP proxy and the population proxy can be expected (see section 2).

Subsets of the dataset were compared using mark permutation tests (Crema et al., 2016; Bevan et al., 2018; Crema and Bevan, 2020), which were done using the `permTest` function. The objective is to compare multiple SPDs to each other rather than a theoretical model. This way the trajectories of different regions, practices or samples can be compared. The null hypothesis of a mark permutation test is that compared SPDs represent samples derived from the same population. Each date is marked with a label (e.g. region A, B; sample type X, Y), and each set is aggregated in a new SPD. Subsequently the labels assigned to the dates are shuffled and each set is again aggregated in an SPD (Crema and Bevan, 2020). This is repeated 500 times to create an envelope against which the subset SPDs can be compared. Subsets of dates from different countries, landscape zones or sample materials were compared to explore biases and identify periods of significant deviation from the norm.

The landscape zones were selected with a European digital elevation model (EU-DEM v1.1, European Environment Agency (EEA): land.copernicus.eu). Three categories were made in QGIS v3.16.8 (<10 m, >10 and <50 m, and >50 m above ordinance level) using the raster calculator, after which the elevation data and categories were extracted and joined using the tool ‘sample raster values’. To visualise how spatial and material subsets of the data overlap, heatmaps were produced per
sample material. These were made by applying heatmap symbology in QGIS using a 100 km radius.

5.2. Model testing frequency distributions

We discuss an overview by climate chronozones of the changing environmental conditions (Table 1, Fig. 3a) and the corresponding density changes in the KDE (Fig. 3d) and model tested SPD results (Fig. 4).

5.2.1. GS2 and the Late Glacial Interstadial / GI-1

The end of GS-2 is characterised by harsh climatic conditions, though in south Germany, conditions already became favourable enough for settlement by Magdalenian groups (Maier, 2015). This is associated with the expansion of steppe and tundra biomes, which gradually spread further north after the sudden pronounced rise in temperatures at the start of GI-1e, opening up new territories for Magdalenian and Hamburgian settlement (Grimm and Weber, 2008; Pettitt and White, 2012; Maier, 2015; Ballin, 2017). The KDE gradually increases throughout these phases (Fig. 3c). The SPD deviates from both models during this phase of initial migration into the region (Fig. 4). This immediately calls attention to the assumptions underlying the models. The exponential model assumes starting levels (0.2) that are not met by the data. This might be because the GS-2 expansions were confined to a limited portion of the research area. Additionally, continuous gradual growth is assumed but already during GI-1e, conditions had ameliorated enough for activity levels surpassing the expectations of this model. The climate model on the other hand reflects the marked increase in δ18O values at the onset of GI-1e, but environmental conditions gradually follow temperature increases, followed in turn by human activity. This explains deviations from the climate model during GS2, and the negative deviation during GI-1e. At the end of GI-1e, the SPD data conforms to the expected envelope of the climate model. From GI-1d onwards, the KDE shows a decline lasting until the start of GI-1c1, after which density increases through GI-1b, peaking during GI-1a. The Allerød decline, while falling within the exponential model’s envelope, is a negative deviation from the climate model. This is surprising, given the temperate conditions and expansions of birch-pine forests in the region and the frequency of Federmesser/Azilian sites associated with the period (Baales and Street, 1996; De Bie and Veermeersch, 1998; Vermeersch, 2011; Pettitt and White, 2012; Riede and Edinborough, 2012; Crombé et al., 2013a; Crombé and Robinson, 2017). During GI-1a, SPD values exceed the climate model, which might indicate that activity of these groups increased towards the end of this period. Additionally the Laacher See eruption may have contributed to large scale burning as well as the preservation of material from preceding periods (Baales and Street, 1996; Street et al., 2012; Riede, 2016).

5.2.2. Younger Dryas / GS-1

GS-1 (Younger Dryas) involved climate deterioration spanning a millennium, considerably affecting soil stability, hydrology and vegetation (Hoek, 1997; Bos et al., 2018). The preceding Laacher See eruption, also had far-reaching effects on the environment and people (Weber et al., 2011; Riede, 2016; Riede and Kierdorf, 2020). The sustained decrease in the dataset may be linked to the comparatively dispersed presence of the Ahrensburgian in large parts of the research area (De Bie and Vermeersch, 1998; Vermeersch, 2011; Weber et al., 2011; Pettitt and White, 2012; Ballin, 2017; Street et al., 2019). During GS-1, the KDE envelope drops almost fully below the lower values of GI-1a. The SPD drops to a negative deviation 1.5 centuries after the onset of GS-1. This delay is visible in the climate model test as positive deviations up until 12.5 ka calBP indicating persistence of activity while conditions deteriorate. At the end of GS-1 activity increases at the same rate as the temperature proxy.

5.2.3. Preboreal

Early Holocene sea level rise (Cohen et al., 2017; Hijma and Cohen, 2019), continuously altered the environmental conditions in our study area. The Early Preboreal marks the return of birch and pine forests, and Ahrensburgian and Federmesser groups dissolving into the Early Mesolithic and its characteristic changes in material culture (Pettitt and White, 2012; Sørensen et al., 2018; Street et al., 2019). After the 11.4 ka event, there is evidence for further innovation in lithic assemblages, characterised by an expanded spectrum of standardized microlithic forms (Street et al., 2019). In terms of activity, the KDE values increase at the end of GS-1, with the envelope rising fully above previous (GS-1) values around 11.4 ka calBP, and dropping directly following the event, recovering a few centuries later. The SPD data barely falls within the envelope of the exponential model test during this event, but in comparison to the climate model, there is a negative deviation around a century after the event, lasting for ca 150 years.

5.2.4. Boreal

Forests diversified during the Boreal with hazel and oak and further sea level rise led to the separation of Britain from the continent around 9 ka calBP, culminating in the final flooding of Doggerland following the 8.2 ka event and the Storegga landslide tsunami (Sturt et al., 2013; Cohen et al., 2017; Gaffney et al., 2020). The KDE fluctuates during the start of the Boreal at 10.8 ka calBP and the 10.3 ka event. The Preboreal to Boreal transition shows a negative deviation from the climate model test, but no significant effect of the 10.3 ka event is observed. Two periods of high density are observed starting ca. 10 ka and 9 ka calBP. Both are significant positive deviations in the exponential model test, but only the Late Boreal peak is significant in the climate model test. The KDE shows a trend of growth peaking at 9.8 ka calBP. This period is followed by a decline until 9.5 ka calBP, which is significant in the climate model test. This sharp decline is not correlated to any patterns in the NGRIP data. However, it lines up with a steep section of the calibration curve (see Fig. 3c), which can indicate cosmic ray or solar storm induced changes in the rate of 14C production in the atmosphere (Reimer et al., 2020; Heaton et al., 2021), both of which could have impacted climate. SPD density increases again to peak around 9 ka calBP and is not impacted by the 9.3 ka event. High density persists for several centuries, followed by decline ca. 8.6–7.7 ka calBP.

The Boreal peaks may be due to migrations related to land loss in Doggerland (esp. around 9 ka), generally favourable environmental conditions in these periods between cold events, or changing cultural factors. The Boreal involves shifts in assemblage types defined by projectile point frequency and technological changes, that coincide with the timing of both peaks (Robinson et al., 2013; Waddington, 2015; Crombé, 2018, 2019; Van Maldegem et al., 2021; Conneller, 2022). However, changes in practices that increase archaeological visibility may also align with such cultural changes (see 5.3).

5.2.5. The 8.2 ka event

The 8.2 ka event marks the end of the Early Holocene (Rasmussen et al., 2014) and the event takes place in the middle of the decline in the KDE and SPD that sets in around 8.6 ka calBP. Climate conditions were regionally harsher during 8.2 ka and the event coincided with the Storegga slide tsunami which variably impacted the coastal zones in the research area (Weninger et al., 2006; Waddington and Wicks, 2017; Griffiths and Robinson, 2018; Blankholm, 2020; Gaffney et al., 2020; Mithen and Wicks, 2021), as well as a sea level jump (Hijma and Cohen, 2019). The KDE declines following the 8.2 ka event, but this trend started centuries earlier indicating the role of other factors. The negative deviation from both model tests starts a century after the 8.2 ka event, indicating these coinciding external events did impact human activity to some extent. Another shift in material culture happened in this period when trapezium arrowheads became more dominant after ca. 8.5–8.3 ka calBP (Robinson et al., 2013; Crombé, 2019).
5.3. Effects of biasing factors

Biases play a considerable role in dataset formation, influencing radiocarbon density. Below, differences between landscape zones, nations, and different sample materials are discussed. Differences can indicate real regional differences in the past, but temporal patterns may also be caused by research focus, policy, preservation conditions or accessibility. Some examples of biases that influence significant patterns in the overall density distribution are detailed below, but the discussion is not exhaustive.

5.3.1. Landscape zones

Ward and Larcombe (2021) argue differences between landscapes in terms of preservation and environmental history obscure underlying patterns when data are aggregated into the same SPD. Climate-driven environmental dynamics such as sea level rise, related preservation conditions and accessibility differ between landscapes at different elevation levels (ibid.). We divided the research area in a low (<10 m above OL), middle (10–50 m above OL) and high zone (>50 m above OL), each zone with a similar sample size (Fig. 6). Permutation tests show how temporal patterns in these different landscape zones (Fig. 7) deviate significantly from the space-averaged temporal patterns presented earlier (4.2, 5.2).

Due to a combination of accessibility, preservation conditions of sediments and settlement history, data density is higher in the higher landscape zones (>50 m) during GS-2 and GI-1e. This is related to the prevalence of Late Magdalenian sites, which are most prominently found in caves and open air sites in this landscape zone (Pettitt and White, 2012; Maier, 2015). Activity in middle and lower zones is underrepresented during this time. The increase into average levels in these zones during GI-1 lines up with the expansion of Hamburghian (Grimm and Weber, 2008) and Federmesser/Azilian groups into these landscapes (De Bie and Vermeersch, 1998; Vermeersch, 2011; Crombé and Robinson, 2017). Simultaneously, activity decreases to average levels in the high zone, possibly because of deteriorating preservation conditions in the Allerød (see 5.3.3). The decrease at the end of GI-1 in the low zone is contemporaneous with increased activity in middle zone of the landscape. Throughout GS-1 and the Preboreal, activity remains above average here and below average in the lower zone, possibly indicating landscape preferences of Ahrensburgian and Early Mesolithic groups (Crombé et al., 2011; Vermeersch, 2011; Street et al., 2019). While climate could be a factor, the overrepresentation of the Dutch dataset plays a considerable role here, with 40 % of sites from this period and zone situated in the eastern and southern Netherlands (see SI-8).

Throughout the Early Boreal (starting ca. 10.7 ka calBP) prevalence of activity shifts towards the lower landscape zone. Surrounding the 10.3 ka event, activity fluctuates in the middle and low landscape zones. The Early Boreal peak appears to shift from the middle (ca. 10.2–9.8 ka calBP) to the low zone (9.8–9.5 ka calBP), while activity is significantly lower in the high zone (9.9–9.4 ka calBP). Similarly, the Late Boreal peak is prominent in the middle zone (9–8.6 ka calBP) followed by low zone activity increasing to above average levels later (8.6–8.4 ka calBP) while simultaneous activity in high zones is below average, dropping further at 8.6 ka calBP. These Boreal peaks might align with drowning Doggerland landscapes west of Belgium and north of the Netherlands (Sturt et al., 2013), but more detailed modelling is required to investigate the timing and extent of these inundations. Following the 8.2 ka event, activity rises in the low zones, fluctuates below average in the middle zone and stays below average in the higher zones. Late Mesolithic sites in the Netherlands account for 70 % of the dates from the lower zone in this period (SI-8), indicating that research bias and changed accessibility of densely populated coastal contexts play a role
The middle landscape may show the effects more reliably, with a significant drop at 8 ka calBP, while activity in higher zones continue to drop from 8.3 to 7.8 ka calBP.

Comparing different landscape zones highlights biases in the data as well as past geographic shifts in density. Providing that biases are considered, potential migrations or geographic clustering can be identified, indicating changing landscape preferences and settlement patterns. These are most prominently visible surrounding the start and end of GI-1 and with the shifting peaks in the Early and Late Boreal. Below we further explore differences and biases on a national scale.

Fig. 7. Permutation test of the defined landscape zones.

Fig. 8. Permutation test comparison between countries.
5.3.2. Countries

Testing for differences between national datasets can highlight biases caused by research focus or national policy (Fig. 8). However, some patterns in the SPDs appear to relate to the timing of expansion into new areas. During GS-2 and GI-1e the SPDs follow a familiar pattern from Germany (>16 ka) to Belgium (15.5 ka) to Britain (15 ka) following the timing of Magdalenian and later Hamburgian settlement in Denmark (14.3 ka peak) and the Netherlands (14 ka peak) (Grimm and Weber, 2008; Pettitt and White, 2012; Maier, 2015). Similar differences in the timing of increased density towards the end of GS-1 (12.2 ka in the Netherlands, 12 ka in Germany, Belgium and Britain and 11.8 ka in Denmark) may indicate regional Ahrensburgian expansions (Vermeersch, 2011; Weber et al., 2011; Street et al., 2019), possibly related to climate conditions.

For the Holocene, the Boreal peaks are clearly different on the national scale. Both peaks are present in Britain, with the first peak significantly more pronounced (10.6–9.7 ka calBP). The appearance of narrow-blade microblade sites in the area between 10.4 and 9.5 ka calBP has been suggested as evidence for migrations out of Doggerland (Waddington, 2015). In Belgium, the Early Boreal peak is also significant (10.2–9.7 ka calBP) and coincides with the appearance of sites with triangle microliths, which may be related to similar responses. Late Boreal activity is much lower in Belgium. In the Netherlands, the Early Boreal peak is absent, with density below average (10.7–9.9 ka calBP), but the Late Boreal peak is prominent and significant (9–8.4 ka calBP).

In Germany, the first peak is also absent, but the data for the whole period after 10.3 ka calBP is below average, apart from the second peak which rises to the average value (9–8.7 ka calBP). In contrast, this period is significantly below average in Denmark (8.8–8.4 ka calBP). The timing of the second Boreal peak in these areas coincides with Middle Mesolithic changes in material culture, such as the appearance of surface retouched points (Crombé, 2019). As seen in 5.3.1, after the 8.2 k event, density increases in the Netherlands and Denmark, while Germany is below average. While past population movements out of Doggerland may explain some of these differences, preservation, accessibility and research focus probably play a considerable role. National differences...
are further discussed below, in the light of preservation and selection of different sample materials.

5.3.3. Sample material

The most common sampled materials are animal remains (bone, antler, teeth), charcoal and plant remains. The preservation of each is dependent on the conditions of their context. Sample materials are also potentially biased in different ways, e.g. old wood effect for charcoal samples, and contamination by organic conservation materials for animal remains. When these effects are multiplied in a large heterogeneous dataset, SPD patterns in periods with predominantly charcoal dates may skew older while patterns in periods with predominantly animal remains may skew younger. Permutation testing highlights differences in the use of sample materials for different periods (Fig. 9) and Kernel

Fig. 10. Geographic distribution of the dataset by sample material, presented in heatmaps. top left: bar chart of sample material frequency per country; top right: heatmap of full dataset; middle and bottom: heatmaps per material category.
Density maps show the spatial distribution (Fig. 10).

Charcoal is less frequently used to date Late Glacial and Preboreal sites compared to archaeology from the Boreal onwards. The inverse goes for animal remains, being used frequently to date Late Glacial sites (except in the Allerød: GI-1-c1 – GI-1a), while animal remains are apparently less available for dating younger sites (<10.2 ka calBP). Plant remains were used more frequently to date Allerød, Preboreal, and Early Boreal sites (up to ca. 9.5 ka calBP). Dates on other materials consist mainly of the dating of sediments and layers related to archaeological sites, which are less prominently used for GS-2/GI-1-e sites and for sites dating between 10.5 and 9 ka calBP. More frequently such dates are used on sites from the second half of the Allerød, GS-1, and between 8.5 and 7.5 ka calBP, possibly reflecting dates on Laacher See tephras, the Usselo layer, younger Dryas coversand and tsunami deposits. To what extent these sampling differences reflect preservation or sampling strategy is difficult to judge.

Preservation bias is dependent on regional differences in geology and lithostratigraphy and influences research bias (Ward and Larcombe, 2021). These biases lead to national differences in used sample material (Fig. 10). Animal remains are a more prominent sample material in the Ardennes, Germany and Britain where better preservation is available in chalky soils and cave deposits. Where animal remains (especially worked bone or antler) are preserved they may be favoured for sampling over charcoal and plant remains. Differences in sampling procedure may play a role here as well (e.g. whether flotation is used or not).

The Netherlands and Belgium (Flanders in particular) are generally overrepresented in our dataset. The dated material predominantly consists of charcoal and plant remains. The Boreal peaks in the Belgian and Dutch subsets may relate to different research strategies for the Mesolithic. The Early Boreal peak is most prominent in the plant remains subset. The preference for hazelnut shells as short-lived samples (Cromb et al., 2013b) may have skewed results towards the timing of the initial spread of hazel (Van Maldegem et al., 2021). Hazelnut accounts for 37% of the Belgian dates in this period, and only for 6% of Dutch dates (SI-8). The Late Boreal peak is only visible in the charcoal subset and coincides with the prominence of pit hearth use in the region. Charcoal from these highly visible features are often the only source for sampling Mesolithic sites on acidic sandy soils (Peeters and Niehus, 2017), and indeed ca. 58% of Dutch dates from this period are from hearths (SI-8). Differing opinions about the nature of hearth pits (Crombe et al., 2015; Huismann et al., 2019, 2020; Crombe and Langohr, 2020) may then in part explain the national differences in the prominence of the Early and Late Boreal peaks, as only 9% of Belgian Late Boreal dates are from hearth pits (SI-8).

6. Discussion and conclusions

Dates-as-data approaches give reproducible and testable results. Proxies are indirect measures of the processes we are interested in and dealing with biases and uncertainty is a fundamental part of their use (Timpson et al., 2015). Instead of dismissing the use of dates-as-data approaches because of biases inherent in our datasets (Becerra-Valdivia et al., 2020; Ward and Larcombe, 2021), we argue for critical explorative analysis of biases as a means of identifying patterns and their underlying causes using permutation testing and the analysis of subsets. We do not think bias can be reduced meaningfully with vetting. Our vetting affected different parts of the dataset disproportionately and timing of changes in the SPDs remained mostly unaffected (see 4.2, SI-6). As vetting introduces a new methodological bias to the dataset, we suggest always comparing the spatial and temporal distribution of full and vetted datasets to explore differences. Rather we suggest an inclusive approach to data collection to improve the sample size of the dataset and potential subsets, and subsequent consideration and analysis of subsets that are potentially biased in different ways.

Radiocarbon density is influenced by a range of factors (see section 2 and 3). Fluctuations in this proxy could be caused by past changes in 1) behaviour that affects population size or density, 2) behaviours with a higher or lower archaeological visibility; or spatiotemporal differences in 3) accessibility, 4) preservation conditions, 5) research focus; or finally 6) the calibration curve. Considering these factors, observed patterns can sometimes be interpreted in multiple, conflicting ways. Climate and environmental change may cause or correlate with past changes in behaviour (factors 1, 2), but related changes such as sea level rise can influence accessibility (3) and preservation conditions (4), while shifts in the calibration curve and climate change can be caused by the same processes (6).

6.1. Potential patterns

Patterns of climate-correlated changes in past behaviour can be identified in the results. In the KDE (Fig. 3d) we observed that the general trend of the dataset is growth and decline corresponding to regime changes in climate, i.e. the transitions from the Late Pleniglacial (GS-2) to the BA-Interstadial (GI-1) to the Younger Dryas (GS-1) to the Early Holocene. Similar adaptive modes coinciding with the climate regimes of the Late Glacial and Early Holocene were established for southwest Europe (Barton et al., 2019). In northwest Europe, these trends line up with known migrations in and out of the research area, population changes and changes in material culture (5.2). Cold phases during the Late Glacial and Early Holocene climate sometimes correspond to dips or flattening of the growth in our dataset notably during GS1 and following the 11.4 ka and 8.2 ka events. Periods of higher density in our dataset line up with more temperate phases, notably the BA-Interstadial and the Boreal. Model testing the SPD and permutation testing of subsets allows for the identification of significant periods of increased and decreased activity. Model tests allow doing this in comparison to hypothesised models of growth, while permutation tests allow for identification of significant differences between subsets. The following specific density changes could be interpreted in terms of human behaviour correlating to climate and environment:

- Gradual expansion into the research area while climate ameliorates at the end of GS-2 and throughout GI-e while temperatures rise quickly (5.2.1), the spatiotemporal differentiation of which is observed by analysing regional subsets (5.3.1, 5.3.2).
- Increased activity at the end of the BA-Interstadial with a relatively high persistence during the transition to GS-1 (5.2.1, 5.2.2). Throughout GS-1 activity is lowered significantly (5.2.2), but it increases at the pace of climate amelioration towards the start of the Holocene (5.2.3), spatial differences in the timing of which can also be tracked (5.3.2). Overall activity gradually increases at the same pace as the temperature proxy (Fig. 5b) but is decreased or increased following climatic, environmental, or cultural changes (5.2.3).
- Decreased activity is observed following the 11.4 ka event and the transition to the Boreal (10.8 ka calBP). Following inundations in Doggerland (10.2–9.6 ka calBP) activity increases, which coincides with material culture changes in some areas, as well as the spread and prevalence of hazel (5.2.4, 5.3.2, 5.3.3).
- The cultural transition to the Middle Mesolithic (9.5 ka calBP) coincides with decreased activity. Further inundations in Doggerland and cultural changes around (9–8.6 ka calBP) correlate to increased activity (5.2.4, 5.3.2), but coincide with the prevalence of the highly visible practice of hearth pit use (5.3.3).
- Finally, overall activity is decreased following the 8.2 ka event (5.2.5), but is regionally variable with a significant increase in lower landscape zones, calling the negative impact of this event
and the contemporaneous Storegga slide tsunami into question for these coastal regions (5.3.1).

6.2. Potential limitations

Of the above, several changes could be explained in part by accessibility, preservation and research focus (5.3), but it is not possible to parse the contribution of past behaviour versus bias to radiocarbon density. Possibly biased sections of our dataset that we identified include the following issues.

- High density during GS-2/GI-1e versus lower density during the Allerød could relate to decreased preservation and accessibility of animal remains for this period (5.3.1, 5.3.3).
- Density during GS-1 may be influenced to a large extent by sites from the Netherlands which are overrepresented in the middle landscape zone (5.3.1).
- The Early and Late Boreal peaks may indicate real changes in past practices, but their prominence is partly caused by research bias and visibility (5.3.3).
- Increased activity in lower landscape zones following the 8.2 ka could be caused by accessibility and research bias, as the Dutch dataset is overrepresented for this period and sea level rise made coastal sites more accessible than for previous periods (5.3.1).
- Overall, the imbalanced spatiotemporal distribution in sample materials (Fig. 8, Fig. 10, SI-8) raises questions, as preservation conditions are both temporally and geographically differentiated (Ward and Larcombe, 2021). Summed dates for periods with different sample prevalence can be skewed in different ways because of such biases.
- Finally, decreased activity around 9.5 ka calBP might be an artefact of the calibration curve (see Fig. 3c, Fig. 5b). Solar storms and cosmic rays influence $^{14}$C production in the atmosphere as well as climate (Heaton et al., 2021), possibly compounding the effect on SPDs. We should therefore remain cautious about fluctuations even when using model tested SPDs. Other areas on the calibration curve that show similar steepness (and related spiking in normalised SPDs, see Fig. 3c) are at ca. 12.75, 12.55, 10.2, 9.0 and 8.4 ka calBP.

6.3. Closing remarks

Analysis of large radiocarbon datasets can highlight the correlation of climate and environmental changes with changes in human activity. This requires a sufficiently large spatiotemporal scale is necessary to ensure an appropriate sample size, as well as an understanding of how different past behaviours can influence radiocarbon date density. Exploring which other underlying factors can influence pattern formation at this macroscale showed that disentangling the effects of biases and past human behaviour remains a challenge. Responsible dates-as-data approaches require an understanding of the archaeology, the timing of external events, the impact of the calibration curve, and of the ways biases may have influenced the formation of patterns in the result.

To move towards a better understanding of patterns caused by biases on the one hand, and past human activity on the other hand, expanding the analysis in relation to landscape dynamics is essential (cf. Ward and Larcombe, 2021), as is closer analysis of the timing of cultural changes (Grimm and Weber, 2008; Vermeersch, 2011; Robinson et al., 2013; Cramb, 2019; Van Maldegem et al., 2021; Conneller, 2022). Both aspects were subject to strong geographical and temporal differentiation throughout, and their understanding is key to elucidating the impact and scope of climate and environmental change on past human societies.

CRediT authorship contribution statement

P.W. Hoebe: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft. J.H.M. Peeters: Conceptualization, Methodology, Supervision, Validation, Writing – review & editing. S. Arnoldussen: Conceptualization, Methodology, Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

supplementary data are available

Acknowledgements

We thank and commend open-data projects whose data we used in this study: ADS, CalPal, EuroEvol, PACEA, RADON and Palaeolithic Europe. Specifically, we want to thank researchers who generously helped us in our search for more available radiocarbon data, namely Christopher Bronk Ramsey, Matthieu Boudin, Marcel Niekus, Bernard Weninger, Erwin Cziesla, Birgit Gehlen and Annabel Zander. We thank Dan Lawrence for exchanging ideas about the methodology and Theo van Mourik (GfT Groningen) for his assistance with programming. We are especially grateful to Michael Dee (GfT Groningen) for supplying data, giving valuable direction and comments on earlier drafts of this paper. Finally we thank Dr Martin Hilz and three anonymous reviewers whose constructive suggestions helped improve the paper.

Funding

This study is part of the PhD research by the first author, financed by the Graduate School of Humanities of the University of Groningen.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jasrep.2023.103944, and https://github.com/pirhoebe.

References


Crombe, P., Robinson, E., Van Strydonck, M., Boudin, M., 2013b. Radiocarbon dating of mesolithic open-air sites in the caverns area of the north-west european plain: