



Research papers

An ICAR spatial model to reconstruct regional geographic variations in arrival times

Alexes K. Mes^{a,*}, Enrico R. Crema^{a,b}^a Department of Archaeology, University of Cambridge, Cambridge, UK^b McDonald Institute for Archaeological Research, University of Cambridge, Cambridge, UK

ARTICLE INFO

Keywords:

Archaeological dispersal
Statistical modelling

ABSTRACT

Reconstructing the speed and pattern of dispersals of cultural traits from the archaeological record is typically based on spatio-temporal distributions of dated artefacts. However, archaeological sampling biases often result in uneven datasets where some regions or time periods are well represented, while others are not. This hinders our ability to offer precise and accurate reconstructions of past dispersal processes. Here we are proposing an approach to address these issues, by introducing a new method to take into account local geographic variations in past movements and refine temporal estimates of regional arrival times by implementing a hierarchical Bayesian Intrinsic Conditional Auto-Regressive (ICAR) model that partially informs the estimated arrival time in a focal region by the estimated arrival times in neighbouring regions. The hierarchical Bayesian ICAR model makes two major advances over previous approaches. Firstly, it provides estimates of the arrival time for areas with little or no archaeological data. Secondly, the model is more flexible than existing phase models or regression analyses, as subjective hypotheses and assumptions about the origin and route of dispersal are not built in. Through a series of experimental applications on simulated archaeological datasets with different dispersal dynamics, we examine the behaviour, accuracy, and precision of the ICAR model against commonly applied phase models and the model's robustness to variation in sample size and unevenness in sampling intensity of archaeological datasets.

1. Introduction

A core challenge in modelling dispersals from the archaeological record is determining when the movement of people or cultural traits occurred. Traditionally, archaeologists and historians have relied on simple visualisations to characterise past dispersals — often in the form of maps with large arrows representing single speeds drawn across a continent. These visualisations mislead. The reality in any human movement is far more complex: the units of migration (individuals, families, communities), as well as the localised tempo and direction of dispersal shifted frequently. Cultural traits and subsistence strategies evolved through multiple interactions and across shifting ecological environments that, in turn, shaped movement trajectories. Statistical models that account for local geographic variations in this movement are needed to determine the dispersal pattern and offer more accurate estimates of regional arrival times — revealing, for example, whether a dispersal spread continuously as a single, sustained event; in waves; or through a series of leapfrog movements.

Dispersal events in the archaeological record — whether demic migration, cultural diffusion or a combination thereof — are spatio-temporal processes framed within the context of a particular set of

environmental and social conditions. Determining *when* a dispersal process occurred is the fundamental initial step from which to test causal hypotheses: *Did ecological conditions accelerate or hinder movement through a landscape? Was population pressure a driving force? Is there evidence that an interaction with incumbent communities enabled or delayed the dispersal?* Defining a more complex spatio-temporal understanding of how movement occurred is, therefore, the key archaeologists need to understand the mechanisms of a dispersal (Ammerman and Cavalli-Sforza, 1973, 1984; Cavalli-Sforza, 2002; Hazelwood and Steele, 2004; Fort, 2012, 2023, 2025) and link the context in which a dispersal happened to the nature of the movement itself (Shennan, 2001; Beaudry and Parno, 2013; Rascovan et al., 2019; Betti et al., 2020; Gregorio de Souza et al., 2022).

The quality of the archaeological record complicates dispersal reconstructions and introduces biases into interpretations of past phenomena. Pre- and post-depositional processes act to obscure underlying records of human behaviour; while the dimensionality, resolution, sampling interval, and scope of our collected data massively constrain the types of questions that can be answered about our past (Perreault,

* Corresponding author.

E-mail address: akm73@cam.ac.uk (A.K. Mes).

2019). Sampling bias is one such issue: the archaeological record is spatially uneven. Legacies of colonialism; arbitrary late 19th and early 20th century borders; current or recent civil unrest; unfavourable environmental conditions all contribute to ‘patchwork’ archaeological sampling, which translates to archaeological datasets where some regions are well-represented, while others suffer an extreme dearth of data.

Statistical methodologies can offer potential avenues to account for uneven spatial sampling in archaeological datasets in ways that rectify past biases and allow for further research on under-examined research areas (Banning, 2021; Crema, 2020). Existing statistical modelling techniques used in archaeology to analyse dispersal rates or to gain accurate estimates of the arrival times in a diffusion¹ include regression-based methods (Bocquet-Appel et al., 2009; Gkiasta et al., 2003; Pinhasi et al., 2005; Van Etten and Hijmans, 2010; Silva et al., 2015) and Bayesian phase models (Blackwell and Buck, 2003; Filipović et al., 2020; Long et al., 2017; Leipe et al., 2020; Crema et al., 2022). However, these methods suffer from a number of limitations that result in difficulties reconstructing the movement of past communities and cultural traits – particularly in areas with little to no evidence – from the existing archaeological record to the degree of accuracy required to evaluate hypotheses proposed by archaeologists and historians.

Regression-based analyses to reconstruct dispersals of cultural traits or subsistence strategies have a long history of use in archaeology, beginning with the pioneering work of Edmonson (1961) in the 1960s who used the earliest occurrences of pottery, copper and maize to determine the rate of diffusion of Neolithic traits globally. A similar approach was undertaken by Ammerman and Cavalli-Sforza (1971) in 1971, who used a linear regression on ¹⁴C dates to determine the rate of the Neolithic expansion in Europe.

The basic premise of regression analyses in this context is to determine the rate of spread or diffusion as the slope of the straight line fitted to the relationship between distance from a putative origin point and radiocarbon age of the ‘earliest’ samples of a particular phenomenon at each site. In attempting to achieve this, regression-based analyses suffer from three major limitations:

- (i) The practical choice of how samples are selected to capture the earliest dates is subjective and spatial scale dependent (i.e. *earliest* relative to where?), and filtering protocols are often not specified in analyses Crema et al. (for further discussion see 2022). Further, the analysis is sensitive to how the centre of diffusion or spread is defined, since its location serves as the reference point from which distance and, therefore, speed is estimated.
- (ii) The method is often based on mean or median calibrated dates and does not account for measurement uncertainty.
- (iii) Typically (as in Russell et al., 2014; Isern and Fort, 2019) the use of linear regression yields single dispersal rates, which dramatically oversimplifies the description of the diffusion process by treating it as a uniform spread across the region of study, and failing to acknowledge local complexities where the tempo and direction of the dispersal might have shifted.

When taken together, these limitations obscure local dynamics and make it difficult to determine negative effects resulting from measurement uncertainty, calibration error, or sample size. Some of the shortcomings of the typical regression-based approach can be addressed: for example, rather than modelling the conditional mean as in ordinary least-squares, a quantile regression model can be used to examine earliest arrival dates from the full sample by estimating the predictor-dependent variable relationship at a specified quantile (e.g. 90th or 95th percentile) Koenker (2005) (see archaeological

applications in Van Etten and Hijmans, 2010; Silva et al., 2015; Riris and Silva, 2021; Rotunno and Crema, 2025). In 2022, Crema et al. (2022) proposed introducing a Gaussian Process adaptation to a Bayesian quantile regression framework (GPQR), whereby an additional parameter (modelled as a multivariate normal distribution) accounts for site-level variation in the tempo of the dispersal process, while fully accounting for measurement error. Unfortunately, the computational intensity of the analysis renders this approach unviable on large datasets. Other regression methods which allow for variations in dispersal rate include non-linear regression (Silva et al., 2015) and geographically weighted regression (Gkiasta et al., 2003). These analyses – like all regression methods – remain predicated on knowing the dispersal point of origin, which is often disputed or unknown.

An alternative approach is to abandon regression-based analyses entirely and model the diffusion process using phase models. Initially used to define the chronological ordering of stratigraphic layers within individual sites, phase model analyses of radiocarbon dates have been adapted in regional studies — where, in a spatial context, geographic areal units of analysis are the equivalent of stratigraphic layers. Typically phase models are used to accurately and precisely chronologically delineate cultural phases (Brunner et al., 2020; Crema and Kobayashi, 2020), or to estimate the arrival of cultural traits (Capuzzo and Barceló, 2015) or crops (Leipe et al., 2020; Long et al., 2018; Crema et al., 2022). Phase models can also be used to model the arrival times of a diffusion front. An early work by Blackwell and Buck (2003) shows how this method can be used to calculate arrival times of populations reoccupying northern Europe towards the end of the Late Glacial period. Phase model limitations, however, include:

- (i) the calculations are sensitive to how the spatial regions are defined;
- (ii) not accounting for sample interdependence, which may bias the arrival time estimation; and
- (iii) the implementation of constraints which are often placed on parameters. In a stratigraphic context, constraints are a useful source of additional information; however, in a regional context, specific arrival time ordering is often based on assumed routes, which reduces the model’s flexibility and builds in subjective hypotheses or strong prior knowledge into analyses.

To elaborate on the first point, sensitivity as to how geographic regions in an analysis are defined is a source of statistical bias known as the Modifiable Areal Unit Problem (MAUP).² If the areal units of analysis are too large, spatial aggregation might mean that important local variations in the parameter of interest (e.g. arrival times) are obscured; while areal units that are too small might suffer from smaller sample sizes, in turn leading to lower precision in the estimated parameters (Leipe et al., 2020). As explained by Buzzelli (2020), it is possible to mitigate the effects of MAUP by first developing a strong rationale for the areal units chosen³ and thereafter examining the stability of the data under variations in scale and placement of areal units.

The second limitation of phase models is more complex. Sample interdependence introduces a bias which a non-hierarchical Bayesian phase model is not able to account for.⁴ It is common for prominent archaeological sites to yield multiple radiocarbon dates and, as such, samples could be distributed unevenly over a given region. Contrast

² First described by geographers (Openshaw, 1984; Amrhein, 1995), MAUP refers to instances whereby summary statistics aggregated over defined areal units are dependent on the shape and scale of the aggregation unit.

³ Options include uniform areal units of a specified shape, as well as areal units with boundaries defined at local changes in geographic features (such as mountains, rivers, or dense vegetation) which might act as potential barriers.

⁴ This is akin to complete pooling where all observations (radiocarbon dates) are statistically treated as if they belong to a single population (rather than different sites within a region).

¹ Dispersal rate and arrival time are inverse parameters as speed is inversely proportional to time.

10 samples from 10 different sites with 10 samples from the same site. The former situation will yield an estimation of the arrival time of a cultural variant that is more representative of the arrival over the whole considered region, while the latter is really only indicative of the arrival at the one site (Crema et al., 2022). Recent research has highlighted two solutions to mitigate uneven sampling density and interdependence:

- (a) The sample could be collapsed — keeping the earliest date in the series at each site or, according to Ward and Wilson (1978), combining multiple within-site dates if these dates are close enough in age to belong to single occupation event. This *subsampling* approach is taken by both Russell et al. (2014) and Isern and Fort (2019) in modelling the Bantu spread. Gregorio de Souza et al. (2020) also did this while modelling late Holocene diffusion in South America, as did Fort et al. (2004) examining Palaeolithic advances in Europe.
- (b) Accounting for the hierarchical structure of the data by building this complexity into the phase model. These hierarchical Bayesian phase models were used by Crema et al. (2022) to study the variation in the dispersal of rice agriculture in prehistoric Japan.

While popular, the first solution, (a), is not a healthy methodology as it assumes that the earliest date available for each site is solely reflective of the first arrival event in a given location and that dates from the same within-site context are identical, i.e. that the difference in dates between samples in the same context is entirely a consequence of measurement error. This is a strong assumption and difficult to justify unless the same sample is being dated multiple times.⁵ Having to subset or filter the data a priori is not good data practice as it might destroy local dynamics not visible in the initial stages of data analysis. Further, this reduces the sample size available for analyses and consequently the precision of the parameter estimates. Considering that sample size is often a restricting feature of many statistical application in archaeology, making use of all available information to inform the model offers a more robust course to adopt. Furthermore, if sub-setted or pre-filtered data are then published in aggregated form, they do not conform to FAIR principles, making it difficult for other archaeologists to reuse it.⁶ For these reasons, sub-sampling and filtering should be avoided.

Solution (b) – modifying a non-hierarchical Bayesian phase model to include a hierarchical structure – is a less problematic approach. The model, graphically depicted in fig.(S1), estimates locally the start date at each site, and then examines the distribution of the inferred parameters to estimate the start date of the local region. Both regression and phase models (including Bayesian hierarchical phase models) can be adapted to correctly handle calibrated radiocarbon dates and take measurement uncertainty into account (see Crema et al., 2022) — though, in practice, this is often not accounted for and the issue ignored. However, even when factoring in various mitigations, these methods share a common failing: they are unable to provide sensible parameter estimates in geographic areas which contain no archaeological information. When examining the arrival time of a diffusion front in a focal region, no assumptions other than the prior are placed on areas that do not contain any observations. This is an implicit

⁵ Even the same object where different samples are taken – for example, charcoal remains outside a pot and food residue within – will likely yield different dates.

⁶ The FAIR data principles state that a dataset should be *findable*, *accessible*, *interoperable*, and *reusable* (Wilkinson et al., 2016). Open data access and computation reproducibility in archaeology has gained traction in recent years, with advocates calling for increased adoption of FAIR within ethical frameworks (which allows for the protection of vulnerable data), citing increased research impact and greater partnerships between academia, industries, and communities (Marwick, 2017; Nicholson et al., 2023).

waste of information: in any diffusion process, movement occurs from a spatial position in a preceding time-step and occupies a different spatial position in a subsequent time-step. The local arrival times are thus spatially autocorrelated, although the structure of that autocorrelation could vary. In other words, the resultant arrival times in neighbouring areas that do contain sites can be used to help reconstruct an estimate of the arrival time in a focal region. This is particularly useful for those regions without archaeological sites or radiocarbon date information. The precision of the arrival date estimates in these regions are likely low; but, it is still an improvement on our initial absence of information in the without-samples regions.

Building on this idea, this paper introduces a new method to estimate dispersal arrival times by implementing a hierarchical Bayesian Intrinsic Conditional Auto-Regressive (ICAR) model. At its core, the model partially informs the estimated arrival time a_k in the region k , by the arrival times in its neighbouring regions $a_{\bar{k}}$ for neighbours $\bar{k} = 1, 2, \dots, n$ (where subscript $\bar{k}$ denotes ‘not including k ’). The hierarchical Bayesian ICAR model makes two major advances over previous approaches. Firstly, it yields useable parameter estimates for areas with limited or no archaeological data. Secondly, the model is more flexible, as subjective hypotheses and assumptions (other than spatial autocorrelation) are not pre-built in. More specifically: (i) unlike in regression-based methods, the hierarchical Bayesian ICAR model does not assume a point of origin of the dispersal process; and (ii) the model does not impose strong prior constraints on the dispersal route as is often done in phase model analyses (see Crema et al., 2022 for an example of constrained phase model analyses on rice dispersal in Japan). The origin or route of a dispersal is often unknown or contested (for example, the early- vs. late-split hypotheses of the Bantu Expansion de Maret, 2013), so there is a strong case for a method which assumes neither. Therefore, this method can be useful for testing dispersal hypotheses in cases where the route is unknown or there are multiple points of origin. In addition, the ICAR model can offer higher precision compared to conventional phase models (see Section 3.2) — both in regions where there are limited or no data and in regions where there are. Finally, as per standard Bayesian phase models, the proposed approach does fully account for measurement uncertainty and calibration errors.

The structure of the model is described in Section 2. This paper then examines the behaviour, accuracy, and precision of the ICAR model against the conventional phase model used in archaeology (Section 3.2) and the robustness of the inferential model to variation in overall sampling intensity and its evenness across space (Section 3.3). To test the inferential power of the proposed ICAR model, this paper uses simulated data rather than an archaeological case study, as with simulated datasets the underlying dispersal patterns and parameters are known, ensuring model performance can be systematically evaluated and compared.

2. Materials and methods

2.1. The hierarchical Bayesian ICAR model

Building on an existing phase model structure, a hierarchical Bayesian ICAR model to describe the calendar date ($\theta_{i,j,k}$) and subsequent observed radiocarbon age ($X_{i,j,k}$) of sample i from site j situated in region k , can be defined as

$$\theta_{i,j,k} \sim \text{Unif}(\alpha_{j,k}, \alpha_{j,k} + \delta_{j,k}), \quad (1)$$

$$\alpha_{j,k} \sim \text{Unif}(a_k, b_k), \quad (2)$$

$$X_{i,j,k} \sim \text{Normal}(f(\theta_{i,j,k}), \sigma_{i,j,k}), \quad (3)$$

where $\alpha_{j,k}$ is the arrival time of the phenomenon at site j in region k ; a_k is the overall arrival time of a given trait in region k ; the term

$(\alpha_{j,k} + \delta_{j,k})$ is the end date at site j in region k ($\delta_{j,k}$ is understood as the duration of occupation); and b_k is the overall end date in region k . Sample interdependence is taken into account by the hierarchical structure introduced in Eqs. (1), (2), where regional arrival times, a_k , are inferred from the start date of individual sites rather than directly from individual calendar dates. Sites with a large number of dates contribute more information than sites with fewer dates, predominately by increasing the accuracy of $\alpha_{j,k}$. The true calendar date of the sample is given by $\theta_{i,j,k}$; while $f(\theta_{i,j,k})$ is its corresponding calibrated ^{14}C age (using the appropriate calibration curve). Measurement error associated with $\theta_{i,j,k}$ is accounted for by Eq. (3), where $\sigma_{i,j,k}$ is the square root of the sum of the squares of the corresponding error in the calibration curve and the sample's radiocarbon age error. The observed radiocarbon age of the sample is then given by $X_{i,j,k}$.

The ICAR distribution is then used to model the spatial autocorrelation in the arrival time a_k in Eq. (2). In modelling diffusion curves (how the probability of a sample, such as a seed, being associated with an innovation, like farming, changes with time), Crema et al. (2024) have used a temporally informed ICAR model. Time is discretised and parameter estimations of the sample's probability, p_t , are conditioned on previous and subsequent values of the parameter, p_{t-1} and p_{t+1} respectively. However, spatially using ICAR models – such that a_k is conditional to its spatial neighbourhood rather than abutting time slices – have not, to our knowledge, been previously used to infer the geographical dispersal rates from the archaeological record; hence the need to carefully outline the model's structure.

By discretising space into geographic areas, parameter values in focal regions are conditionally informed by parameter values in their neighbouring regions. ICAR distributions serve as improper priors⁷ for random fields, such as temporal or spatial processes. The prior is defined jointly over a set of latent process values across the entire parameter space (Besag, 1974; Rue and Held, 2005). Mathematically, this is defined using conditional distributions, where each observation is modelled as a normal distribution centred on a weighted average of neighbour observations (Besag, 1974; Rue and Held, 2005). In our model,

$$p(a_k | a_{\bar{k}}, \tau) \sim \text{Normal}\left(\frac{1}{\omega_{\bar{k}}} \sum_{m \in N_{\bar{k}}} \omega_{km} a_m, \omega_{\bar{k}} \tau\right), \quad (4)$$

for

$$\omega_{\bar{k}} = \sum_{m \in N_{\bar{k}}} \omega_{km}, \quad (5)$$

where the arrival time a_k at a location k is also conditionally dependent on the parameter values of its adjacent, regional neighbours, $a_{\bar{k}}$. In Eq. (4), the term $n_{\bar{k}}$ is the number of neighbouring regions to k , $N_{\bar{k}}$ is the set of these neighbours; and ω_{km} is the unnormalised weight associated with a pair of adjacent locations, k and m , where $m \in N_{\bar{k}}$. The scalar precision of the spatial random effects is controlled by τ .⁸ The *sum-to-zero* constraint is not implemented, i.e. $\sum_k a_k \neq 0$, which allows the spatial field to shift up or down globally.

The end parameter, b_k , is modelled as an uniform distribution (for example, $b_k \sim \text{Unif}(50, 5000)$) and is not of interest in reconstructing the arrival of a given trait or cultural package into a region.⁹ Considering the tendency among archaeologists to sample and record earlier dates in the period under study (for example, the over-sampling of Early

Iron Age material compared to Later Iron Age material within an archaeological site in sub-Saharan Africa Clist et al., 2023), as well as the propensity of some traits to evolve (rather than disappear) over time, it is more difficult to use such archaeological databases to reconstruct the absolute 'end' of a trait in a region.¹⁰

2.2. Bayesian areal wombling to identify uneven dispersal rates

In studies of cultural dispersal processes, spatial variation in the speed of spread can reveal important clues about the processes shaping that dispersal. Areas of rapid change in dispersal speed of cultural traits often indicate the influence of exogenous effects or interactions which alter how quickly a cultural trait is transmitted. Rapid change in speed might arise, for example, in areas where ecological conditions shift, requiring adjustments in subsistence strategies or settlement patterns. Or, perhaps, swift change in trait dispersal might indicate an area of active contact between incoming groups and incumbent communities — where processes such as trade or negotiation can modify the pace and direction of cultural transmission. These regions often give rise to dynamic cultural identities and fluid socio-economic and political relationships. Identifying the spatial boundaries where such changes in dispersal speed occur in a continuous diffusion process is therefore of archaeological interest.

William Womble developed a method for this type of boundary analysis in 1951, in order to detect geographic boundaries between populations with continuous variation in genetic or morphological traits (Womble, 1951). Wombling methods have since been used in biostatistics and public health (Jacquez and Greiling, 2003; Liang et al., 2009; Lu and Carlin, 2005; Ma et al., 2006, 2010), as well as ecology (Fitzpatrick et al., 2010; Wheeler and Waller, 2008) to identify regions of considerable flux. These methods show great promise for use in archaeology and are presented here used alongside an ICAR approach (see the previous Section 2.1) in order to comprehensively reconstruct past movement and highlight major accelerations and decelerations in the dispersal. Following Banerjee et al. (2015), Bayesian areal wombling assigns a boundary likelihood value (BLV) to an edge shared between a pair of neighbouring regions (k, m):

$$\Delta_{km} = |a_k - a_m|. \quad (6)$$

The BLV posterior can be estimated concurrently with the ICAR arrival time during a Markov chain Monte Carlo MCMC analysis. Fuzzy wombling¹¹ can then be used to determine the probability that the boundary of (k, m) is important in relation to a chosen base-line, ζ — i.e. the probability that the absolute difference in dispersal time between two adjacent regions is higher than ζ . Given G total samples drawn during a MCMC analysis, boundary importance is calculated as:

$$\hat{p}_{km} = \frac{1}{G} \sum_{g=1}^G \mathbb{I}[\Delta_{km}^{(g)} > \zeta], \quad \text{where } \Delta_{km}^{(g)} = |a_k^{(g)} - a_m^{(g)}|, \quad (7)$$

and g represents a single MCMC iteration. The specified threshold is given by ζ and is chosen depending on the range and scale of the analysis. For example, in Section 3.2, a diffusion process which takes place over 1500 BP (from around 4700 BP to 3200 BP) is simulated. When the threshold is chosen as $\zeta = 500$ years, the areal wombling method described in Eqs. (6), (7) will then highlight the probability that the diffusion process between these adjacent regions took more than 500 years. Through wombling, a numerical continuous variable

⁷ In Bayesian analyses, improper priors are not valid probability distributions, since they do not integrate to a finite value over the entire parameter space. They are useful, however, when together with a likelihood function, they yield a proper posterior distribution.

⁸ A high τ implies stronger spatial autocorrelation (i.e. where a_k are constrained to be similar across neighbouring regions), while a low τ implies a weaker spatial dependency.

⁹ In some sense, b_k can be assumed to be identical for all regions (fixed) — marking the end point of chronological sampling in the dataset.

¹⁰ Projects looking at the extinction rate of cultural traits and how particular practices are abandoned are possible and of archaeological relevance, but fall outside the remit of this paper.

¹¹ As opposed to Crisp wombling where a value of 1 is assigned to the boundary of (k, m) if and only if the BLV exceeds the threshold ζ (Fitzpatrick et al., 2010; Banerjee et al., 2015).

is translated into a probabilistic assignment to two categorical levels. Archaeologists can use this metric to determine unusually high or low differences in arrival times between adjacent regions, providing, in the former case, a visual approach to identify where potential cultural or environmental barriers in the diffusion process might have been.

While Eqs. (6), (7) describe mean-based wombling, residual-based wombling can also be performed (Banerjee et al., 2015). In this extension, the arrival time in the ICAR model (Eq. (4)) is adapted to include linear additions of weighted covariates and an ICAR distribution which models the unaccounted effects, i.e. the spatial residuals, ϕ_k . Using ϕ_k instead to calculate the BLVs identifies boundaries that remain unexplained by the modelled covariates.

Because both mean- and residual-based boundaries are estimated from spatially discrete areas, their location and significance could depend on how the sample region is partitioned.¹² Therefore, in a rigorous analysis of an archaeological diffusion process, the resolution and orientation of the chosen spatial grid should be varied – and the wombling analysis repeated – in order to ensure that the detected boundaries are not artifacts of the imposed spatial structure, but reflect genuine areas of rapid flux in the data.

2.3. Sample region and simulated datasets

In order to test the model's performance, a series of simulations were carried out within a sample region, roughly the size of a bounding box covering the region, i.e. $\sim 19,000,000 \text{ km}^2$. In Fig. 1, the sample region is divided into smaller, hexagonal areas for analyses. Hexagonal grids are commonly used in spatial statistical analyses as they offer several advantages over square or rectangular grids. The most important of these in the implementation of the ICAR model is a greater spatial isotropy: hexagonal grids have six equidistant neighbours, while square grids create unequal neighbour distances (considering edge and corner adjacency¹³) which potentially introduces bias in neighbourhood relationships. Hexagonal grid structures therefore minimise edge effects, as well as better approximate spatial processes that diffuse isotropically (Birch et al., 2007).

The simulation window was divided into 81 hexagonal areas, each with a diameter of 600 km^2 . When examining real archaeological data, the definition of the areal units should maximise the number of areas with site information while maintaining a spatial resolution that can capture local patterns. As such, the size of the hexagonal regions can be chosen at the threshold at which decreasing the spatial unit size stops yielding a greater proportion of areas with site information.

This paper will present a series of *tactical simulations* — the term was first employed in an archaeological context by Orton (1973) and refers to the practice of testing whether proposed archaeological methods are able to recover underlying patterns from a simulated (and, therefore, known) reality (Lake, 2014). The sample region in Fig. 1 is used to simulate archaeological data and test the hierarchical ICAR model's ability to infer the parameter of interest, i.e. the arrival time of a dispersal process, a_k , for each spatial unit. In the simulations described in Sections 3.1, 3.2, structured datasets of 800 sites across the sample region were created. A clustering parameter, c , was introduced to control the degree of spatial clustering (see Fig. 7), mimicking the sampling density variation typically found in archaeological datasets where spatial coverage reflects aggregated settlement patterns, survey biases based on individual choices¹⁴ and landscape accessibility, political and

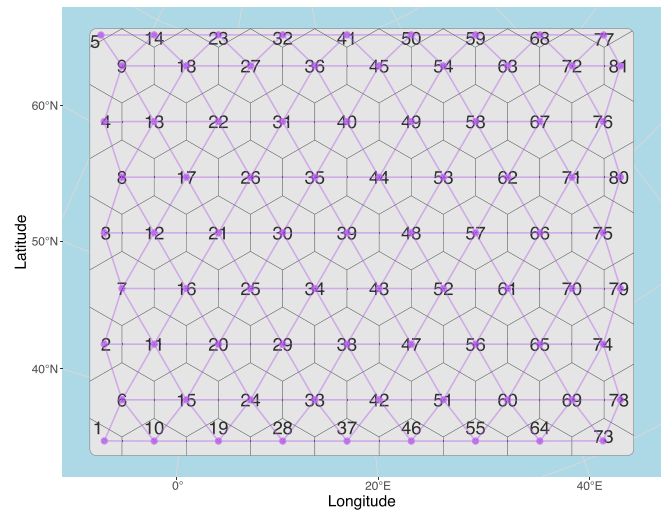


Fig. 1. The sample window used to simulate archaeological data divided into 81 local, hexagonal areas. The area centres and the Delaunay triangulation connecting these centres are depicted. The projected co-ordinate system is ETRS89-LAEA (EPSG code: 3035), used for across-Europe statistical mapping.

historical contexts, as well as preservation and environmental biases. For $c = 0$, a binomial point process generates a spatially random geographical site distribution; while increasing $c \rightarrow 1$ sites are distributed according to the Thomas point process (Thomas, 1949; Baddeley et al., 2016) and the site pattern across the region gradually becomes more clustered (i.e. the sampling density bias becomes more pronounced). A minor clustering bias is introduced in these simulated datasets ($c = 0.3$) while the effect of varying c is explored in Section 3.3.¹⁵

Across each of these 800 sites, 2400 dates (three times the number of sites) are allocated using weighted random sampling from a normal distribution producing varying date observations per site. Dates are then back-calibrated in ^{14}C with an associated error of 20 years. Calendar date values, θ_{ijk} , are assigned by simulating the arrival times, a_k , according to an underlying diffusion pattern. In Section 3.1 arrival times are generated according to a simplified Wave of Advance (WoA) (Ammerman and Cavalli-Sforza, 1971) structure:

$$a_k \sim \text{Normal}(\mu_k, \sigma), \text{ where } \mu_k = \beta_0 - d_k/s. \quad (8)$$

Eq. (8) describes a diffusive process, whereby movement occurs radially from a putative origin at a uniform velocity. Distance from the centroid of the origin hexagon is given by d_k , β_0 is the offset (i.e. the start of the dispersal in the origin hexagon), and s is the speed of the wavefront (in km/year). In Sections 3.2, 3.3, arrival times are generated using a proper conditional autoregressive (CAR) prior (ϕ_k) with additional covariates contributing to the arrival time in each hexagonal region, such that the response μ_k becomes

$$\mu_k = \phi_k + \beta_1 x_{1k} + \beta_2 x_{2k}, \quad (9)$$

where β_1 and β_2 determine the magnitude of each covariate (in *years*), while x_{1k} and x_{2k} indicate the presence/absence of the covariate in the hexagonal area k . Eq. (9) captures a generic diffusion-like process with variable directions of movement conditioned by external covariates that nonetheless yield spatially autocorrelated arrival time. Two covariates were included in Eq. (9) as an example — there is no intrinsic restriction to the number of covariates added. It is worth noting that precisely how the data is simulated is not particularly relevant. The underlying spatial process generating the data — whether Eq. (8) or Eq. (9)

¹² This is another example of MAUP (see Section 1).

¹³ This is known as *Queen adjacency* which allows diagonal influence resulting in eight neighbours to each square area. The more commonly used structure of *Rook adjacency* assumes only edge adjacency, resulting in four neighbours to each area.

¹⁴ Crema and Bevan (2021) discuss how uneven archaeological attention from teams or individual archaeologists may result in site, phase, or areal over-representation.

¹⁵ Varying the size of the database is also investigated in this subsection.

– and that used in the model to infer the parameter of interest (Eq. (4)) are different. This is in order to demonstrate the flexibility of the ICAR model to correctly infer parameters under different scenarios, as when working with actual archaeological data the underlying diffusion process is unknown and complex — the summation of many different effects and interactions.

In all simulated datasets, site arrival times ($\alpha_{j,k}$) were then modelled according to Eq. (2) as an uniform distribution, while the site duration parameter ($\delta_{j,k}$) was randomly sampled from a gamma distribution with a shape parameter of 5 and rate parameter of 0.04 — leading to durations clustered around a mean of 125 years, typically in the range of 70–180 years. In Section 3.2 which considers measurement uncertainty, calendar dates were sampled according to Eq. (1), then back-calibrated in ^{14}C age and assigned a measurement error of 20 years.

2.4. Parameter inference

For each simulated dataset, the inference was performed using the *nimble* v.1.0.1 (de Valpine et al., 2020) and *nimbleCarbon* v.0.2.1 (Crema, 2022) packages in R v.4.3.0 (Team, 2020) to obtain posterior estimates of the model parameters (specifically, the local arrival times a_k are of interest here). MCMC sampling was implemented with the following settings: 2 000 000 iterations, with 1 000 000 dedicated to burn-in; thinning with parameters sampled every 100 steps; four chains; and the Metropolis–Hastings adaptive random walk sampler for all parameters with *adaptInterval* of 5 000 and *adaptFactorExponent* of 0.1. Trace plots and the Gelman–Rubin (R-hat) diagnostic¹⁶ (Gelman and Rubin, 1992) were used to evaluate chain convergence; and can be found for all simulations in the Supplementary Material.

3. Testing the model

3.1. Wave of advance simulation

A Wave of Advance simulation was first performed in order to determine whether the ICAR model was able to recover arrival times in a simple diffusive process. A simulated dataset of 800 sites and 2400 associated calendar dates were created according to Eq. (8). The dataset describes a diffusion process which spanned roughly 1710 years — beginning around 6450 BP and expanding radially from this putative origin until the simulated trait or group arrives in the last area around 4740 BP. Although the ICAR model is capable of handling measurement uncertainty and calibration errors (see Eq. (3)), in this example calendar dates were used in the simulation and the inferential process in order to set a base-line for the model's performance.

Visual inspections of Figs. 2(a) and 2(b) suggests that the ICAR model can successfully recover the correct arrival time pattern of the diffusion process. The performance of the model can be quantified in terms of accuracy¹⁷ and precision¹⁸: it recovers the simulated a_k values with a 95% HPDI accuracy of 96.3%, and achieves a precision of 196 years.¹⁹

¹⁶ The Gelman–Rubin (R-hat) diagnostic compares the variance within each MCMC chain to the variance between the chains. If the chains have converged their variances should be similar and the resultant R-hat will approximately be equal to 1.

¹⁷ Accuracy is defined here as the number of true simulated arrival times that fall within the respective inferred credible intervals (HPDI) divided by total number of simulated local arrival times, i.e. 81.

¹⁸ Precision is defined as the average width of the 95% credible interval. Higher precision is given by narrower intervals.

¹⁹ This is approximately 11% of the expansion range of 1710 years.

3.2. Comparing the ICAR model to a phase model

In the previous WoA simulation, the underlying diffusion process used to simulate the arrival times was a simple radial expansion from a putative origin at a constant speed. The dispersal process can be more complex: in Fig. 3(a) the arrival times have been generated according to Eq. (9) with $\beta_1 = -500$ years affecting areas 1, 2, 6, 10, 11, 15, 19, 20, 24, 25, 29, and 38. The simulated dispersal process began around 2400 BP and arrived in the last area circa 790 years — spanning roughly 1610 years.

The ICAR model's accuracy (95% HPDI) at recovering the complex simulated arrival time pattern is 93.8% with an average precision of 350 years²⁰ (see Fig. 3(b)). Considering this simulation is more complex — with (i) the use of radiocarbon dates (with associated measurement uncertainty) rather than calendar dates and (ii) a simulated arrival pattern including covariates — the ICAR model's poorer precision compared to the WoA simulation (Section 3.1) is expected.

Fig. 3(d) highlights boundaries where the diffusion process has been delayed by 500 years or more, using the fuzzy, mean-based wombling method (Eqs. (6), (7)). A Delaunay triangulation was used to connect the centres of neighbouring areas (see Fig. 1) — creating 208 boundaries shared between the 81 hexagonal areas to consider. Inspecting figure (S7) and Fig. 3(d) simultaneously, the delay caused by the β_1 covariate along the top boundary of areas 2, 11, 20, 29, and 38, has been picked up in the inference process and is highlighted as prominent. Additional boundaries along areas 58 and 63, along 71 and 75, as well as 74 arising stochastically from the simulated dispersal process — and not from the presence of the β_1 covariate — are also emphasised as causing delay.

The performance of this paper's proposed ICAR model at recovering the arrival times in a dispersal process is compared against the hierarchical Bayesian phase model²¹ (discussed in the Introduction 1). In implementing an hierarchical Bayesian phase model, an uniform prior replaces Eq. (2) in modelling the arrival time, a_k . The ICAR and phase models are tested against one another on the same simulated dataset (see Fig. 3(a)). The phase model recovers the a_k 's with an accuracy of (95% HPDI): 96.3%, but with an average precision of 602 years — 1.7 times worse than the ICAR model.

The poor performance of the hierarchical Bayesian phase model is largely due to its inability to provide informed posteriors in regions where no site information exists.²² Fig. 4 demonstrates this clearly²³ — in the 'site-less' regions 32, 50, 59, and 68 the broad posteriors from the phase model indicate that the model is able to learn very little and returns the prior in these regions. The ICAR model, in comparison, uses site information from neighbouring regions to infer the local arrival times. The ICAR model's posteriors in region 32, 50, 59, and 68 are still broad, but they represent a marked improvement on the existing phase model.

There are rare example regions where the phase model is able to infer the arrival time with slightly lower precision and improved

²⁰ This precision is approximately 22% of the expansion duration 1610 years.

²¹ A non-hierarchical, or complete pooling, Bayesian phase model is more common in archaeological applications. However, Crema et al. (2022) showed the hierarchical Bayesian phase model is more precise than the non-hierarchical Bayesian phase model when recovering arrival times in a simulated dispersal process. Therefore, to evaluate the performance of our proposed ICAR model, we compare it against the more precise hierarchical approach.

²² When site information does exist, the ICAR model still outperforms or equals the precision of the phase model. Only considering areas with samples, the precision of the ICAR model is 291 years on average, while the phase model has a precision of 454 years.

²³ For the ICAR model and phasemodel posteriors for all areas see figures (S9–S10).

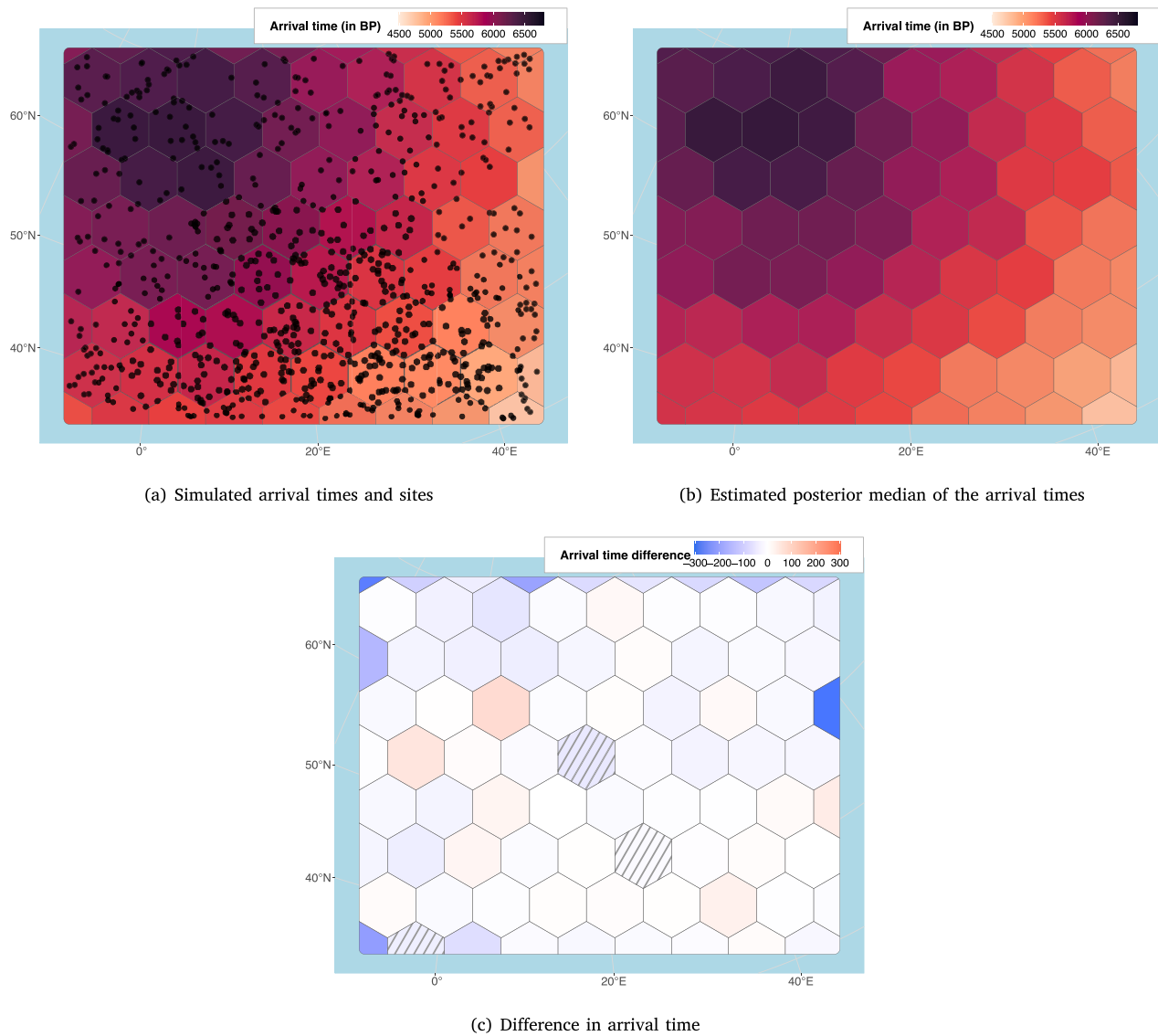


Fig. 2. A Wave of advance (WoA) simulation where figure (a) depicts the 800 simulated sites and the arrival times in each hexagonal region; and figure (b) shows the posterior median arrival time in each region as inferred by the ICAR model described in Section 2.1. Figure (c) depicts the difference in arrival time between the simulated arrival time and the posterior median arrival time in each hexagonal area. Hatching indicates areas where the simulated arrival time does not fall within 95% HPDI of the inferred arrival time posterior. The model's accuracy (95% HPDI) at recovering the simulated arrival times is 96.3% with an average precision of 196 years. Calendar dates without associated errors are used in this toy example.

accuracy. For example, in region 24 the spatial smoothing effect in the ICAR model is potentially too strong — and the arrival times in neighbouring regions (four of which, regions 15, 19, 20, and 29 have an additional influence from the presence of the β_1 covariate) are biasing the arrival time estimate in region 24. The strength of the spatial autocorrelation in the ICAR distribution (τ in Eq. (4)) can be tuned during the MCMC analysis — with slightly lower τ (weaker smoothing) offering potentially better inference in this simulation. Alternatively, the model can be extended: currently τ is assumed to be stationary over the entire sample region, but the degree of spatial auto-correlation might vary across space. Some regions may be more influenced by their neighbours than other regions. Allowing τ to vary over space — i.e. introducing non-stationarity — could improve the inference, although this might not be worth the cost in terms of computational efficiency.

Overall, the ICAR model performs better at inferring the arrival time parameters than the phase model in regions with no or little site information, which is particularly useful for archaeological datasets with geographic biases in sampling efforts (see Introduction (1)). The

relationship between precision and number of sites in a region for both models is given in Fig. 5. For small numbers of sites ($n < 10$), the ICAR model is clearly superior; while for large numbers of sites ($n > 10$), the ICAR and phase model are approximately equally precise.

3.3. Robustness to variation in sample size and sampling intensity

3.3.1. Sample size

The analysis performed in both this and the following Section 3.3.2, assesses how well the ICAR model can recover the underlying dispersal pattern given simulated datasets with differing conditions (here, varying number of sites and in Section 3.3.2, the degree of clustering). In order to examine the effect of overall sample size, 10 datasets were simulated in the same sample region (Fig. 1) with the number of sites, n , varying from 200 to 2000. The number of calendar dates is given by $n_d = 3n$, and as such, varies from 600 to 6000. The clustering parameter remains consistent, $c = 0.5$.

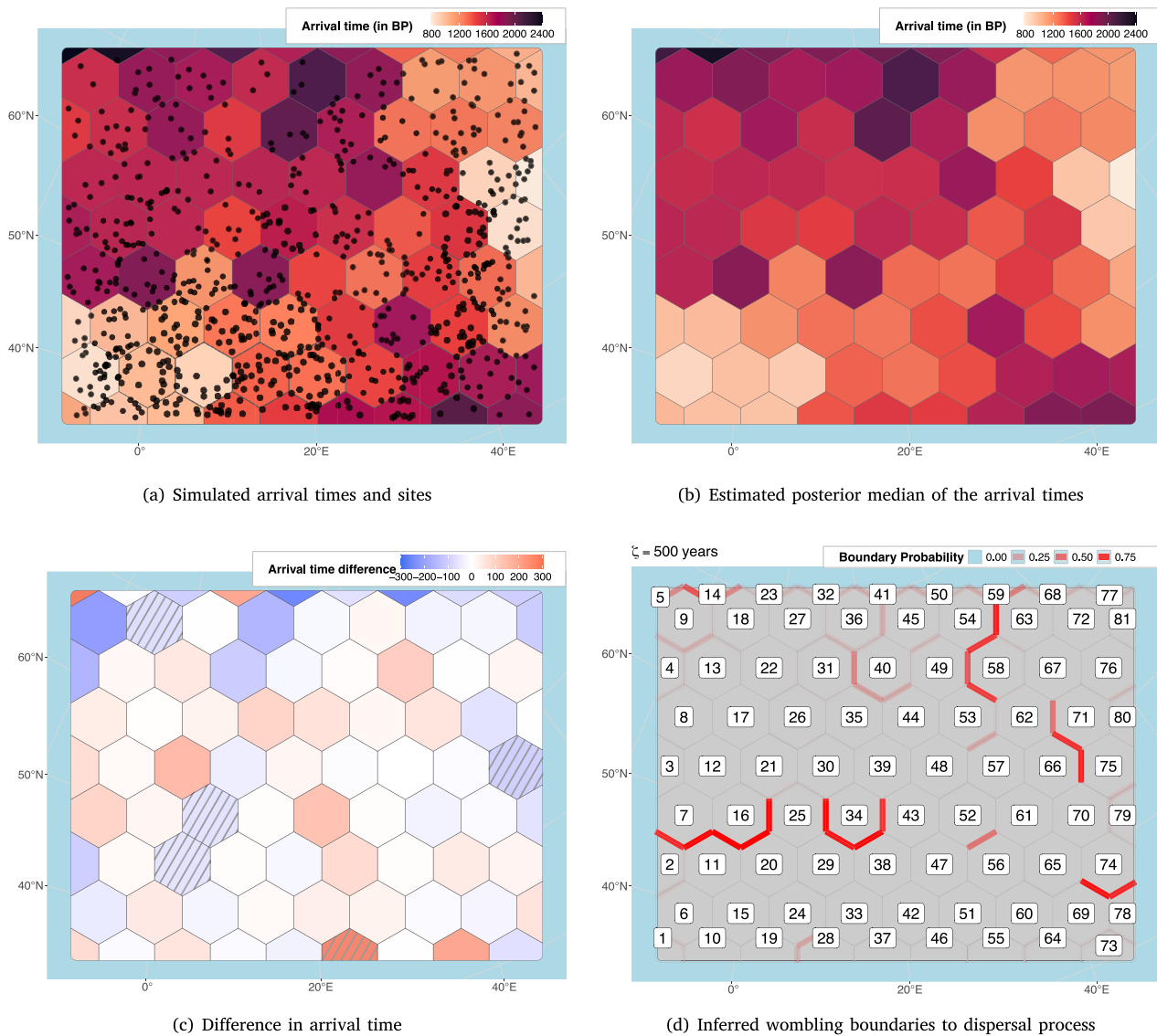


Fig. 3. Using the ICAR model to reconstruct the local arrival times in a simulated diffusion process. Figure (a) depicts a dataset of 800 simulated sites and arrival times in each hexagonal region. The inferred results are given in figures (b) and (d): specifically, figure (b) presents the inferred arrival times from the ICAR model (see Section 2.1); and figure (d) depicts important boundaries over a threshold of $\zeta = 500$ years, highlighted according to a fuzzy, mean-based wombling method (Eqs. (6), (7)). High boundary probabilities – i.e. where a high proportion of the posterior distribution of the absolute difference in dispersal time between two adjacent regions is greater than ζ – are indicated by darker shades of red. Figure (c) depicts the difference in arrival time between the simulated arrival time and the posterior median arrival time in each hexagonal area. Hatching indicates areas where the simulated arrival time does not fall within 95% HPDI of the inferred arrival time posterior. Radiocarbon dates calibrated on the IntCal20 (Reimer et al., 2020) curve with an associated error of 20 years are used in this example.

The robustness of ICAR to variation in sample size is shown in Fig. 6. The model's accuracy remains consistently high – ranging between 85% and 100% – while the average precision improves with larger datasets. More specifically, in Fig. 6(b) there is a threshold between 400 – 600 sites, across which a dramatic improvement in the model's precision is observed. When $n \gg 600$ sites, precision does not improve markedly. This can be helpful for archaeologists intending to use this ICAR method to under the diffusion or expansion process in their region of interest. Knowing how many sites their archaeological dataset consists of, they can use Fig. 6(b) to see what precision they can roughly expect and therefore whether the use of this ICAR method is appropriate for their context. Obviously, it is not the total number of sites which is important in their considerations, but rather the total number given the size of their region of interest. Here, the sample region (Fig. 1) is approximately 19,000,000 km², and as such 600 sites

per 19,000,000 km² (or 1 site per 31,000 km²) can roughly be considered a benchmark for reasonable precision performance from the ICAR model. Luckily, most archaeological datasets do not suffer from such an extreme paucity of data.

3.3.2. Sample intensity

As geographic unevenness in sampling efforts is a common problem in archaeological data (discussed in the Introduction (1)), it is useful to examine the model's performance under increasing degrees of spatial clustering in the site distribution. Examining the effect of the spatial variation in sampling intensity on the accuracy and precision of the ICAR model's inference required the creation of 10 datasets of consistent site number (800 sites) and varying clustering degree, $c = n/10$ where $n \in \mathbb{Z}$, $0 \leq n \leq 10$. As $c \rightarrow 1$, the sampling-density bias in the dataset becomes more pronounced (see Fig. 7).

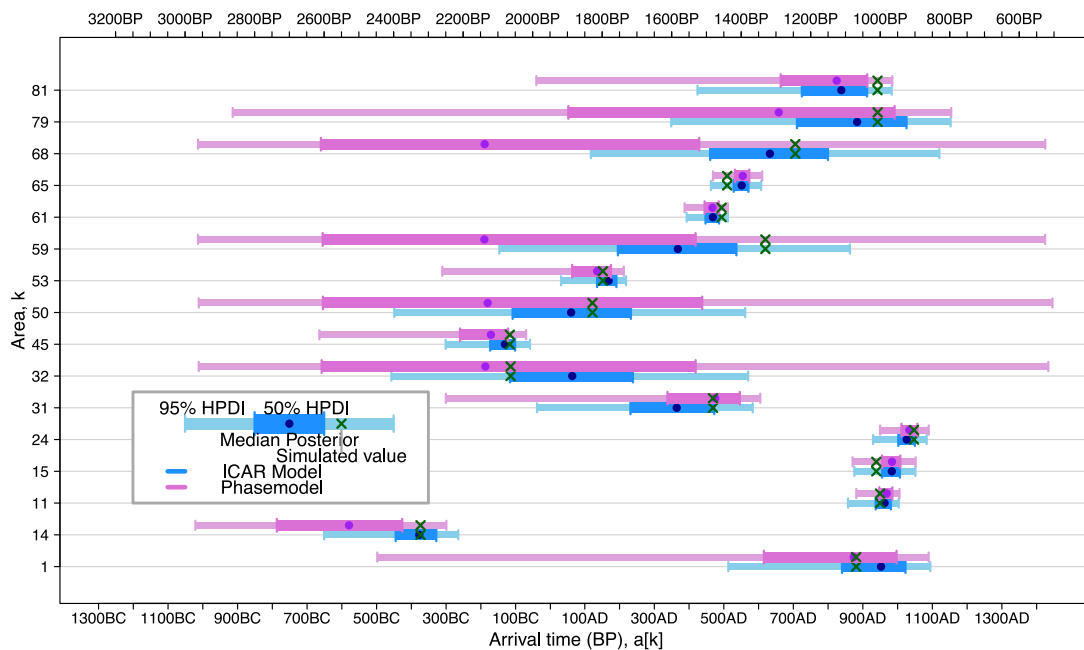


Fig. 4. Posterior distributions of arrival times, a_k , in selected regions k , using the ICAR model (blue) and phase model (purple). The true arrival date in each region is given by a green X.

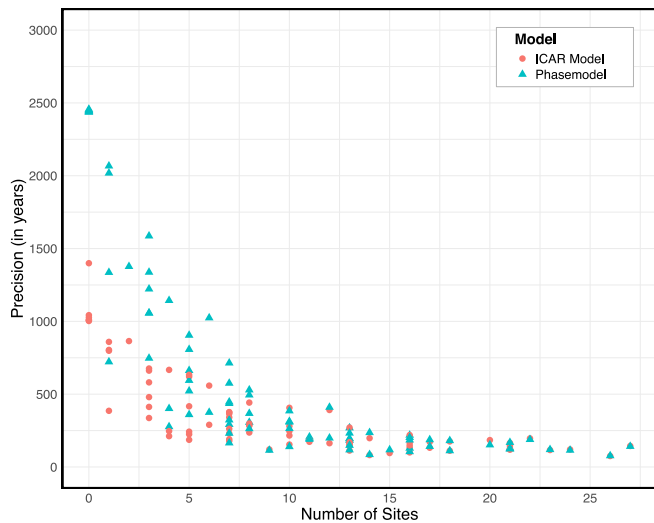


Fig. 5. The relationship between model precision and number of sites in each local hexagonal area for the ICAR model (red, circle) and the phase model (blue, triangle).

The ICAR model's performance on these 10 simulated datasets is depicted in Fig. 8. The accuracy of the model hovers steadily between 85% and 100%, while the precision worsens considerably in more highly clustered datasets. For datasets with large c , there are many areas without sites and many with an over-abundance. While the ICAR model is able to infer parameter estimates from neighbours when site information in a local area is absent or low, this inference has a broader posterior than if the area has some sites (see Fig. 4), and hence the inference precision worsens. In Fig. 8(b), this can be seen around $c = 0.7$ and $c = 0.8$, where the precision increases from 328 ± 343 years in the former case, to 537 ± 475 years in the latter.

4. Discussion

Thinking in terms of neat *bounded cultures* completing singular, linear movements across large areas is rarely useful. Dispersal processes of cultural traits involves, to a lesser or greater extent, human movement. It is inherently complex, with changing units of dispersal (individuals, families, communities) and continually shifting local speeds. During the dispersal in situations where interaction occurs, agency for social action needs to be afforded to both the incumbents in a given area and the migrants — both having an equal ability to be influenced by the other to adopt appropriate cultural traits (Robb and Miracle, 2007).

For example, consider the micro-dynamics of subsistence strategies during a dispersal process. In the late 1970s, John Alexander proposed the concept of a *food production frontier* to explain the transition between a hunter-gatherer-fisher lifestyle and one predominantly focused on livestock or agriculture in Neolithic Europe (Alexander, 1977, 1978). A range of possible frontiers demarcating differences in subsistence strategies captures the dynamism and varied nature of interactions that could occur. The frontier can move and is fluid — at times a rigorously defended, unyielding barrier and at other times malleable and porous. The nature of the frontier dictates the types of relations between the parties with different subsistence strategies. To capture these nuances, the first step is to identify where they might be occurring by establishing the local movement dynamics of the dispersal process.

Building on the structure of commonly implemented Bayesian phase models, this paper introduces a novel approach to infer arrival times in past dispersals by implementing a hierarchical Bayesian ICAR approach which partially informs the estimated arrival time in a region by the arrival times in its neighbouring regions. To our knowledge, ICAR models have not previously been applied spatially to investigate diffusion dynamics in the archaeological record. The model succeeds in improving two major shortcomings of existing phase model and regression-based analyses: primarily, it creates useable parameter estimates for areas with no or limited archaeological data, which aims to improve inference where data are heavily affected by spatial sampling biases inherent to the archaeological record. Additionally, the model does not require or impose strong prior constraints and assumptions on

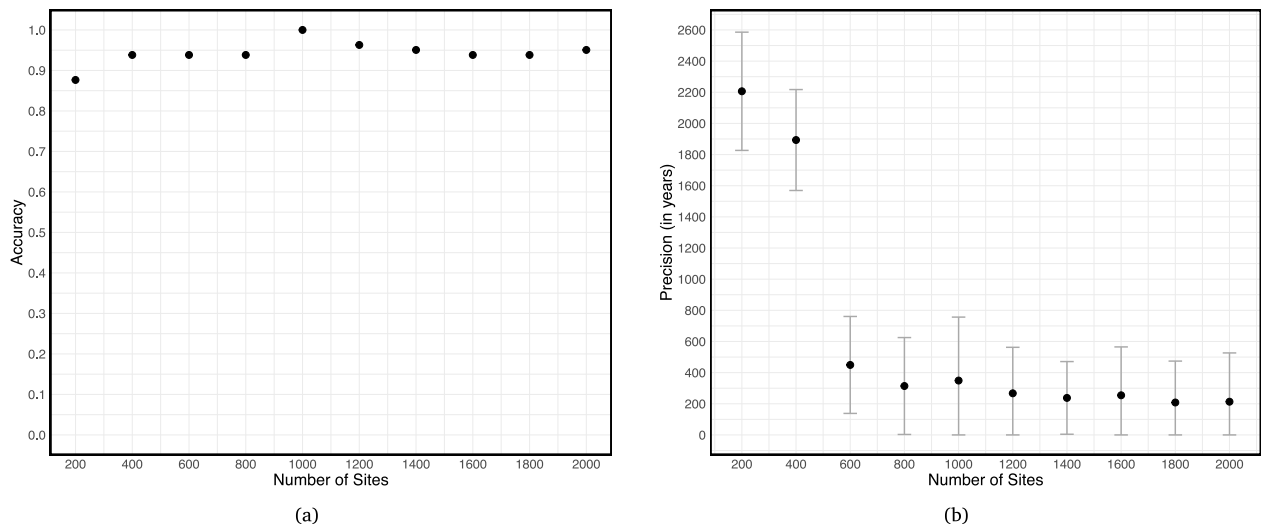


Fig. 6. ICAR model performance measured over ten simulations in which the total number of sites increased from 200 to 2000. Figure (a) shows the accuracy of simulations with each number of sites. Figure (b) presents the average precision for each number of sites, with error bars indicating one standard deviation of precision across the 81 local hexagonal areas.

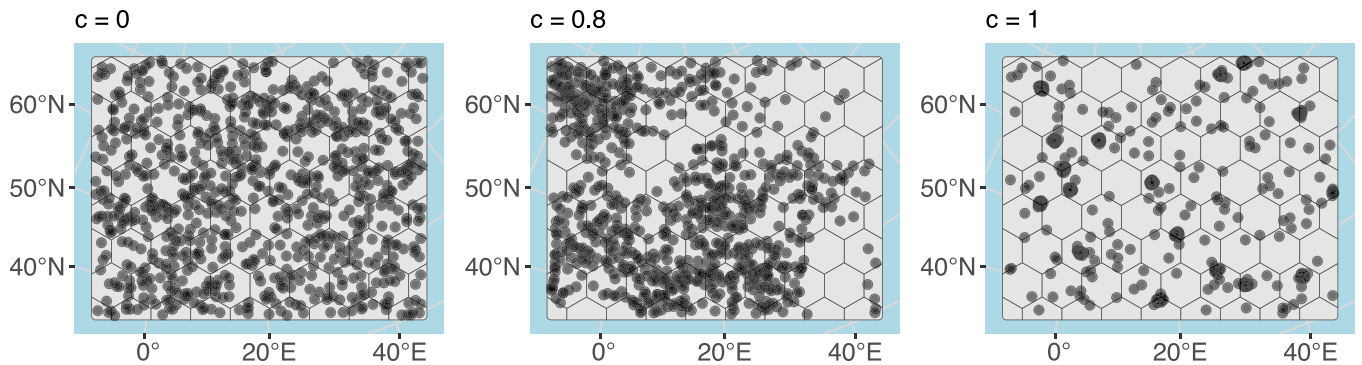


Fig. 7. Increasing the spatial variation in sampling intensity of the simulated datasets by adjusting the degree of data clustering, c . For $c = 0$, a binomial point process produces a spatially random distribution of sites, while increasing $c \rightarrow 1$ sites are generated according to a Thomas point process (Thomas, 1949; Baddeley et al., 2016), resulting in increasingly clustered spatial patterns and a more pronounced sampling-density bias.

the dispersal route and origin. This is particularly helpful when multiple, competing hypotheses are present — for example, the early-split and late-split hypotheses of Bantu expansion movement in sub-Saharan African beginning around 4000 BP (Phillipson, 1977; Vansina, 1990; Huffman, 2007; De Filippo et al., 2012; Currie et al., 2013; Grollemund et al., 2015). The movement of cultural traits across landscapes is inherently complex: bifurcations and inversions in movement pathways, as well as the potential of multiple origins, are all plausible when reconstructing dispersal processes. Consequently, archaeologists require models that are sufficiently flexible to accommodate these complexities.

Through a series of tactical simulations mimicking archaeological datasets (Orton, 1973), we first demonstrated the ICAR model's ability to correctly infer the arrival times, a_k , in a simple diffusive process a Wave of Advance simulation with calendar dates. The model performed with a precision approximately 11% of the expansion range. Further, the ICAR model was considerably more precise than its existing phase model counterpart (Section 3.2). On a simulated radiocarbon dataset, the ICAR model inferred the arrival times with an average precision of approximately 22% of the expansion duration; while the existing phase model only achieved 47% precision on average — largely due to this

model's inability to provide accurate estimates in regions with little or no site information. Lastly, the model's ability to remain robust under varying sample sizes and sampling intensity was explored (Section 3.3). As demonstrated in Sections 3.2 and 3.3, our ICAR model was capable of fully accounting for measurement uncertainty and calibration errors.

The presence of calibration plateaus in the radiocarbon record could also potentially affect a model's ability to accurately infer the timing of a dispersal in the archaeological record. One of the most infamous calibration plateaus²⁴ occurs approximately between 800 and 400 BC (Reimer et al., 2020). During this (or any other) calibration plateau, the relationship between radiocarbon ^{14}C dates and calendar years on the calibration curve flattens out — converting ^{14}C dates to calendar dates results in less precision compared with other chronological intervals, assuming measurement conditions and sample quality are otherwise equal. Since, multiple distinct radiocarbon determinations can correspond to overlapping calendar ranges, chronological uncertainty inflates which potentially obscures genuine diffusion patterns.

²⁴ This plateau is sometimes referred to as the Hallstatt Plateau after the Hallstatt culture of Central and Western Europe.

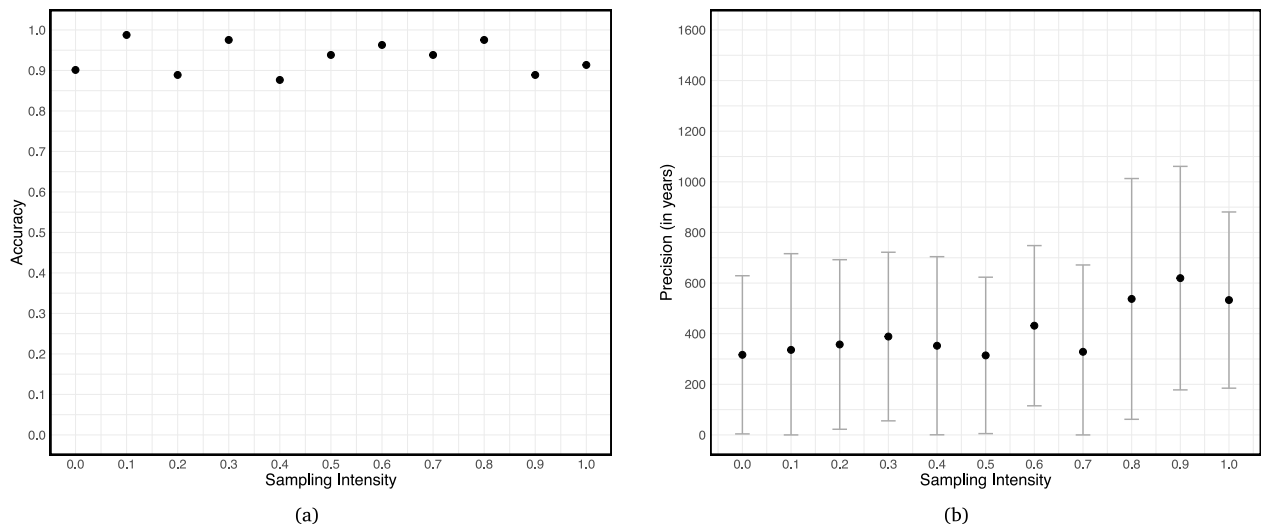


Fig. 8. Measuring the ICAR model's performance over 10 simulated datasets where the sample intensity is varied by increasing the clustering parameter, $c = n/10$ where $n \in \mathbb{Z}$, $0 \leq n \leq 10$. Figure (a) gives the accuracy of simulations at each sampling intensity. Figure (b) depicts the average precision at each sampling intensity, with error bars indicating one standard deviation of precision across the 81 local hexagonal areas.

How the temporal range falls within a calibration plateau – whether the start, end, a section of, or the entire range of the dispersal occurs within a calibration plateau – changes the extent to which the plateau effects the model's convergence. As a result, it is important that individual archaeological practitioners check the convergence of the model within their context-specific cases.

In general, archaeologists seeking to use the hierarchical ICAR model to better under dispersal processes in their specific contexts, should first apply the simulation framework discussed in this paper to examine whether their existing sample structure is sufficient in terms of accuracy and precision. This approach, referred to as *what-if* experiments and first suggested in the context of Bayesian chronological modelling by Christen and Buck (1998), is advantageous as it provides a costless way to examine the likely gains from the data accrual before collection takes place — allowing the practitioner to employ a spatially targeted collection strategy (Buck and Meson, 2015) (for example see Bayliss and Woodman's work in Northern Ireland (Bayliss and Woodman, 2009)).

It is recommended that users of the model carefully consider the scale-dependent nature of dispersal dynamics. As discussed in relation to Modifiable Areal Unit Problem (MAUP) in Section 1, deciding on the size and shape of the areal unit of analysis remains a key consideration and a strong rationale needs to be developed for each context. Potential ideas include (i) choosing a tessellated shape the size of which is determined by the threshold at which decreasing the spatial unit size stops yielding a greater proportion of areas with site information; or (ii) defining areal units by potential barriers (physical factors such as mountains, rivers, or dense vegetation).

In addition, the choice of prior for τ could be further explored. Spatial diffusion processes are, by nature, spatial autocorrelated — the unit of interest originates *from* one location and transitions *to* another in the proceeding time step. In our ICAR model we have assumed a stationary autocorrelation (i.e. τ does not vary across areas in Eq. (4)). Future extensions of this work could consider the effect of varying τ — which would likely better capture more nuances in speed of the diffusion front, but at a computational expense. The method is already fairly computationally heavy: the simulation described in Section 3.2, the ICAR model reconstructing the diffusion process with simulated radiocarbon dates, takes an approximate runtime of 48 h.²⁵ The computational cost depends on the spatial resolution and dataset

size — and will increase with the number of areas and the number of radiocarbon dates.

It is also worth noting that the extent to which the approach proposed here is capable of detecting anomalous movements, such as leapfrog diffusion, is in part dictated by the choice of the prior of τ . A leapfrog movement could entail an arrival date that is significantly earlier than the surrounding regions, effectively countering the assumption of spatial autocorrelation that is assumed in our approach. A non-spatial model might in this case yield estimates signalling the possible presence of a leapfrog with a smaller sample size, while an ICAR model might lead to later arrival time to maintain spatial consistency with neighbouring regions. This tension highlights the trade-off in spatial smoothing approaches: the ICAR model reduces sensitivity to local fluctuations that may arise from sampling variability or measurement error, but this could dampen unusually early, but genuine, signals associated with leapfrog movements. However, as long as the strength of the spatial autocorrelation parameter is not too strong (τ is not too large) genuine early signals will still be apparent in the posterior estimates, particularly when supported by larger number of dates and a stronger effect size (i.e. a substantially earlier arrival date). If the practitioner expects in a particular archaeological case study that leapfrog dynamics play a central role in the dispersal process, additional methods might be required to investigate these movements in detail.

Further extensions to the model can be considered. The ICAR model presented in this paper is tested with simulated radiocarbon dates. In some archaeological contexts, radiocarbon data might only be sparsely available, while other types of chronological information are additionally known. In principle, it is perfectly possible to incorporate other types of chronological evidence. However, the different forms of uncertainties and measurement error for each type would need to probabilistically quantified (Bevan and Crema, 2021). This could be approached similarly to how other Bayesian chronological models integrate forms of chronological evidence other than radiocarbon dates (for example, stratigraphic relationships, historically dated events, material culture typologies) or dates from other chronometric methods, such as dendrochronology. See Buck et al. (1992) and Bayliss (2015) for

²⁵ Estimates of runtime are based on Ubuntu 24.04.3 running on a Dell Inc. PowerEdge R650xs with 128GiB RAM (DIMM DDR4 Synchronous Registered (Buffered) 3200 MHz (0.3 ns)) and 96 Intel(R) Xeon(R) Gold 6336Y CPU @ 2.40 GHz each with 36864 KB cache.

further discussions on this; or Buck et al. (1996) for the Bayesian framework to combine dendrochronological and radiocarbon information. Archaeologically-inferred calendar ages can, therefore, be included alongside radiocarbon dates in the model — as long as these observations have uncertainties quantified such that they can be accounted for in the model. This is harder to do for archaeologically-inferred calendar ages than it is for radiocarbon dates. As such, the integration of multiple lines of chronological evidence will need to be done carefully, on a case-by-case archaeological basis.

In the present implementation of the ICAR model, radiocarbon dates associated with a site or region informing the arrival (a_k) and end (b_k) time of a region are assumed to be uniformly distributed (see Eqs. (1)–(2)). The choice of the uniform distribution represents the activity within a phase with a minimal number of parameters, which allows the behaviour of the hierarchical ICAR model to be demonstrated in a simple form, focussing on the spatial component of the model. Alternatively, other distributions could be used to model the temporal structure of the data. For example, a trapezoidal phase model, allowing for gradual onset and decline, might instead be a more realistic representation of many cultural or demographic dispersals (Lee and Ramsey, 2012; Ramsey, 1995). However, trapezoidal models introduce additional parameters for the durations of the introductory (ξ_k) and declining (δ_k) phases of activity; replacing Eq. (2) such that arrival at a site is drawn from a trapezoidal distribution with four parameters $[a_k, a_k + \xi_k, b_k - \delta_k, b_k]$. The added complexity of estimating four parameters ($a_k, \xi_k, \delta_k, b_k$) for each region, instead of two in the case of the uniform distribution (a_k, b_k) is entirely possible, but requires a strong context-specific justification and sufficient data to identify the parameters reliably. Further, the increased number of parameters could potentially lead to worse convergence in the MCMC analyses. This aside, the hierarchical ICAR model could theoretically be adapted to model the temporal structure of the data with different distributions and this could be explored in future applications.

A third extension to the model could consider adding a formalised treatment of outliers. The presence of outliers are a valid concern with many archaeological datasets, and may arise due to a number of factors. Examples include, measurement error, sample bias in relation to the calibration curve, or the sample not relating to the timing of the event being dated. The practitioner can carefully screen the dataset prior to analysis; identifying and removing problematic outliers. But, it would also be possible to formally account for outliers within the model — potentially by following the approach implemented in OxCal (introduced in Ramsey, 2009), where each date is assigned a prior probability of being inconsistent with its depositional context. Again careful implementation would be required as computational cost might significantly increase with the higher-dimensional parameter space.

Finally, extending the framework to adapt to residual-based wombling, discussed in Section 2.2, would identify boundaries that remain unexplained by modelled factors which influence the diffusion; allowing, through the process of elimination, the identification of regions where decelerations in the dispersal process occur that can be explained by specific covariates.

CRedit authorship contribution statement

Alexes K. Mes: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Enrico R. Crema:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

Code availability

The code required to reproduce the analyses in this paper are available on the GitHub repository (https://github.com/AlexesMes/ICAR_dispersal_model) and archived in the Zenodo repository (<https://zenodo.org/records/20039212>).

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.

Acknowledgements

A. Mes would like to acknowledge the Mandel Rhodes Foundation and the Leverhulme Trust for funding through the Shaun Johnson Memorial Scholarship.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jas.2026.106590>.

References

- Alexander, J., 1977. The 'frontier' concept in prehistory: the end of the moving frontier. *Hunters, Gatherers First Farmers beyond Eur.* 25–40.
- Alexander, J., 1978. Frontier studies and the earliest farmers in europe. *Soc. Organ. Settl.* 13–29.
- Ammerman, A.J., Cavalli-Sforza, L.L., 1971. Measuring the rate of spread of early farming in europe. *Man* 674–688.
- Ammerman, A.J., Cavalli-Sforza, L.L., 1973. A population model for the diffusion of early farming in europe. *Explan. Cult. Chang.* 343–357.
- Ammerman, A.J., Cavalli-Sforza, L.L., 1984. *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton University Press.
- Amrhein, C.G., 1995. Searching for the elusive aggregation effect: evidence from statistical simulations. *Environ. Plan. A* 27 (1), 105–119.
- Baddeley, A., Rubak, E., Turner, R., 2016. *Spatial Point Patterns: Methodology and Applications with R*, vol. 1, CRC press Boca Raton.
- Banerjee, S., Carlin, B.P., Gelfand, A.E., 2015. *Hierarchical Modeling and Analysis for Spatial Data*, second ed. Chapman and Hall/CRC.
- Banning, E.B., 2021. Sampled to death? The rise and fall of probability sampling in archaeology. *Am. Antiq.* 86 (1), 43–60.
- Bayliss, A., 2015. Quality in Bayesian chronological models in archaeology. *World Archaeol.* 47 (4), 677–700.
- Bayliss, A., Woodman, P., 2009. A new bayesian chronology for mesolithic occupation at mount sandel, Northern Ireland. In: *Proceedings of the Prehistoric Society*, vol. 75, Cambridge University Press, pp. 101–123.
- Beaudry, M.C., Parno, T.G., 2013. *Archaeologies of Mobility and Movement*. Springer.
- Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 36 (2), 192–225.
- Betti, L., Beyer, R.M., Jones, E.R., Eriksson, A., Tassi, F., Siska, V., Leonardi, M., Maisano Delser, P., Bentley, L.K., Nigst, P.R., et al., 2020. Climate shaped how neolithic farmers and European hunter-gatherers interacted after a major slowdown from 6,100 BCE to 4,500 BCE. *Nat. Hum. Behav.* 4 (10), 1004–1010.
- Bevan, A., Crema, E.R., 2021. Modifiable reporting unit problems and time series of long-term human activity. *Phil. Trans. R. Soc. B* 376 (1816).
- Birch, C.P., Oom, S.P., Beecham, J.A., 2007. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecol. Model.* 206 (3–4), 347–359.
- Blackwell, P.G., Buck, C.E., 2003. The late glacial human reoccupation of north-western europe: new approaches to space-time modelling. *Antiquity* 77 (296), 232–240.
- Bocquet-Appel, J.-P., Naji, S., Vander Linden, M., Kozłowski, J.K., 2009. Detection of diffusion and contact zones of early farming in europe from the space-time distribution of 14c dates. *J. Archaeol. Sci.* 36 (3), 807–820.
- Brunner, M., von Felten, J., Hinz, M., Hafner, A., 2020. Central European early bronze age chronology revisited: A Bayesian examination of large-scale radiocarbon dating. *PLoS One* 15 (12), e0243719.
- Buck, C.E., Cavanagh, W.G., Litton, C.D., Scott, M., 1996. *Bayesian Approach to Interpreting Archaeological Data*. Wiley Chichester.
- Buck, C.E., Litton, C.D., Smith, A.F., 1992. Calibration of radiocarbon results pertaining to related archaeological events. *J. Archaeol. Sci.* 19 (5), 497–512.
- Buck, C.E., Meson, B., 2015. On being a good Bayesian. *World Archaeol.* 47 (4), 567–584.
- Buzzelli, M., 2020. Modifiable areal unit problem. *Int. Encycl. Hum. Geogr.* 169.
- Capuzzo, G., Barceló, J.A., 2015. Cultural changes in the second millennium BC: a Bayesian examination of radiocarbon evidence from Switzerland and catalonia. *World Archaeol.* 47 (4), 622–641.
- Cavalli-Sforza, L.L., 2002. Demic diffusion as the basic process of human expansions. Examining Farming/Language Dispersal Hypothesis. *Camb.: McDonald Inst. Archaeol. Res.* 79–88.
- Christen, J.A., Buck, C.E., 1998. Sample selection in radiocarbon dating. *J. R. Stat. Soc. Ser. C. Appl. Stat.* 47 (4), 543–557.

- Clist, B., Denbow, J., Lanfranchi, R., 2023. Using the radiocarbon dates of central africa for studying long-term demographic trends of the last 50,000 years: potential and pitfalls. *Azania: Archaeol. Res. Afr.* 1–59, Publisher: Routledge. eprint: <https://doi.org/10.1080/0067270X.2023.2215649>.
- Crema, E.R., 2020. Non-stationarity and local spatial analysis. *Archaeol. Spat. Anal.* 155–168.
- Crema, E.R., 2022. Nimblecarbon: Models and utility functions for Bayesian analyses of radiocarbon dates with NIMBLE (v.0.2.1). R Packag. URL <https://github.com/ercrema/nimbleCarbon>.
- Crema, E.R., Bevan, A., 2021. Inference from large sets of radiocarbon dates: software and methods. *Radiocarbon* 63 (1), 23–39.
- Crema, E.R., Bloxam, A., Stevens, C., Vander Linden, M., 2024. Modelling diffusion of innovation curves using radiocarbon data. *J. Archaeol. Sci.* 165, 105962.
- Crema, E.R., Kobayashi, K., 2020. A multi-proxy inference of jōmon population dynamics using bayesian phase models, residential data, and summed probability distribution of 14c dates. *J. Archaeol. Sci.* 117, 105136.
- Crema, E.R., Stevens, C.J., Shoda, S., 2022. Bayesian analyses of direct radiocarbon dates reveal geographic variations in the rate of rice farming dispersal in prehistoric Japan. *Sci. Adv.* 8 (38), <http://dx.doi.org/10.1126/sciadv.adc9171>, Publisher: American Association for the Advancement of Science. URL <https://www.science.org/doi/full/10.1126/sciadv.adc9171>.
- Currie, T.E., Meade, A., Guillon, M., Mace, R., 2013. Cultural phylogeography of the bantu languages of sub-saharan africa. *Proc. R. Soc. B: Biological Sci.* 280 (1762), Publisher: Royal Society.
- De Filippo, C., Bostoen, K., Stoneking, M., Pakendorf, B., 2012. Bringing together linguistic and genetic evidence to test the bantu expansion. *Proc. R. Soc. B: Biological Sci.* 279 (1741), 3256–3263.
- de Maret, P., 2013. Archaeologies of the bantu expansion. In: Mitchell, P., Lane, P.J. (Eds.), *The Oxford Handbook of African Archaeology*. Oxford University Press, pp. 627–643. <http://dx.doi.org/10.1093/oxfordhb/9780199569885.013.0043>.
- de Valpine, P., Adler, C., Turek, D., Michaud, N., Anderson-Bergman, C., Obermeyer, F., Cortes, C.W., Rodríguez, A., Lang, D.T., Paganin, S., 2020. NIMBLE: MCMC, particle filtering, and programmable hierarchical modeling. 2020. R Packag. Version 0.12 2, URL <https://zenodo.org/record/8047617>.
- Edmonson, M.S., 1961. Neolithic diffusion rates. *Curr. Anthr.* 2 (2), 71–102.
- Filipović, D., Meadows, J., Corso, M.D., Kirleis, W., Alsleben, A., Akeret, Ö., Bittmann, F., Bosi, G., Ciută, B., Dreslerová, D., et al., 2020. New AMS 14c dates track the arrival and spread of broomcorn millet cultivation and agricultural change in prehistoric europe. *Sci. Rep.* 10 (1), 13698.
- Fitzpatrick, M.C., Preisser, E.L., Porter, A., Elkinton, J., Waller, L.A., Carlin, B.P., Ellison, A.M., 2010. Ecological boundary detection using Bayesian areal wombling. *Ecology* 91 (12), 3448–3455.
- Fort, J., 2012. Synthesis between demic and cultural diffusion in the neolithic transition in europe. *Proc. Natl. Acad. Sci.* 109 (46), 18669–18673.
- Fort, J., 2023. Neolithic transitions: Diffusion of people or diffusion of culture? In: *Diffusive Spreading in Nature, Technology and Society*. Springer, pp. 327–346.
- Fort, J., 2025. Tendencies in the tempo of prehistoric agricultural expansions. *J. Archaeol. Res.* 1–49.
- Fort, J., Pujol, T., Cavalli-Sforza, L.L., 2004. Palaeolithic populations and waves of advance. *Camb. Archaeol. J.* 14 (1), 53–61.
- Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statist. Sci.* 7 (4), 457–472.
- Gkiasta, M., Russell, T., Shennan, S., Steele, J., 2003. Neolithic transition in europe: the radiocarbon record revisited. *Antiquity* 77 (295), 45–62.
- Gregorio de Souza, J., Alcaina Mateos, J., Madella, M., 2020. Archaeological expansions in tropical south america during the late holocene: assessing the role of demic diffusion. *PLoS One* 15 (4), e0232367.
- Gregorio de Souza, J., Ruiz-Pérez, J., Lancelotti, C., Madella, M., 2022. Environmental effects on the spread of the neolithic crop package to south Asia. *Plos One* 17 (7), e0268482.
- Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C., Pagel, M., 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proc. Natl. Acad. Sci.* 112, Publisher: Proceedings of the National Academy of Sciences.
- Hazelwood, L., Steele, J., 2004. Spatial dynamics of human dispersals: Constraints on modelling and archaeological validation. *J. Archaeol. Sci.* 31 (6), 669–679.
- Huffman, T., 2007. *Handbook to the Iron Age: the Archaeology of precolonial farming societies in southern Africa*. University of KwaZulu-Natal Press.
- Isern, N., Fort, J., 2019. Assessing the importance of cultural diffusion in the bantu spread into southeastern africa. *PLoS One* 14 (5), e0215573. <http://dx.doi.org/10.1371/journal.pone.0215573>, Publisher: Public Library of Science. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0215573>.
- Jacquez, G.M., Greiling, D.A., 2003. Geographic boundaries in breast, lung and colorectal cancers in relation to exposure to air toxics in long island, new york. *Int. J. Health Geogr.* 2 (1), 4.
- Koenker, R., 2005. *Quantile regression*, vol. 38, Cambridge University Press.
- Lake, M.W., 2014. Trends in archaeological simulation. *J. Archaeol. Method Theory* 21, 258–287.
- Lee, S., Ramsey, C.B., 2012. Development and application of the trapezoidal model for archaeological chronologies. *Radiocarbon* 54 (1), 107–122.
- Leipe, C., Long, T., Wagner, M., Goslar, T., Tarasov, P.E., 2020. The spread of rice to Japan: Insights from Bayesian analysis of direct radiocarbon dates and population dynamics in east Asia. *Quat. Sci. Rev.* 244, 106507.
- Liang, S., Banerjee, S., Carlin, B.P., 2009. Bayesian wombling for spatial point processes. *Biometrics* 65 (4), 1243–1253.
- Long, T., Leipe, C., Jin, G., Wagner, M., Guo, R., Schröder, O., Tarasov, P.E., 2018. The early history of wheat in China from 14c dating and Bayesian chronological modelling. *Nat. Plants* 4 (5), 272–279.
- Long, T., Wagner, M., Tarasov, P.E., 2017. A Bayesian analysis of radiocarbon dates from prehistoric sites in the haidai region, east China, for evaluation of the archaeological chronology. *J. Archaeol. Sci.: Rep.* 12, 81–90.
- Lu, H., Carlin, B.P., 2005. Bayesian areal wombling for geographical boundary analysis. *Geogr. Anal.* 37 (3), 265–285.
- Ma, H., Carlin, B.P., Banerjee, S., 2010. Hierarchical and joint site-edge methods for medicare hospice service region boundary analysis. *Biometrics* 66 (2), 355–364.
- Ma, H., Virnig, B.A., Carlin, B.P., 2006. Spatial methods in areal administrative data analysis. *Ital. J. Public Health* 3 (3–4).
- Marwick, B., 2017. Computational reproducibility in archaeological research: Basic principles and a case study of their implementation. *J. Archaeol. Method Theory* 24 (2), 424–450.
- Nicholson, C., Kansa, S., Gupta, N., Fernandez, R., 2023. Will it ever be fair?: Making archaeological data findable, accessible, interoperable, and reusable. *Adv. Archaeol. Pr.* 11 (1), 63–75.
- Openshaw, S., 1984. The modifiable areal unit problem. *Concepts Tech. Mod. Geogr.*
- Orton, C., 1973. The tactical use of models in archaeology—the SHERD project. *Expl. Cult. Chang.* 137–139.
- Perreault, C., 2019. The quality of the archaeological record. The Quality of the Archaeological Record, University of Chicago Press, <http://dx.doi.org/10.7208/9780226631011>, Publication Title. URL <https://www.degruyter.com/document/doi/10.7208/9780226631011/html>.
- Phillipson, D.W., 1977. *The Later Prehistory of Eastern and Southern Africa*. Africana Pub..
- Pinhasi, R., Fort, J., Ammerman, A.J., 2005. Tracing the origin and spread of agriculture in europe. *PLoS Biol.* 3 (12), e410.
- Ramsey, C.B., 1995. Radiocarbon calibration and analysis of stratigraphy: the oxcal program. *Radiocarbon* 37 (2), 425–430.
- Ramsey, C.B., 2009. Dealing with outliers and offsets in radiocarbon dating. *Radiocarbon* 51 (3), 1023–1045.
- Rascovan, N., Sjögren, K.-G., Kristiansen, K., Nielsen, R., Willerslev, E., Desnues, C., Rasmussen, S., 2019. Emergence and spread of basal lineages of yersinia pestis during the neolithic decline. *Cell* 176 (1), 295–305.
- Reimer, P.J., Austin, W.E., Bard, E., Bayliss, A., et al., 2020. The IntCal20 northern hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* 62 (4), 725–757.
- Riris, P., Silva, F., 2021. Resolution and the detection of cultural dispersals: development and application of spatiotemporal methods in lowland south america. *Humanit. Soc. Sci. Commun.* 8 (1), 1–13.
- Robb, J., Miracle, P., 2007. Beyond migration versus acculturation: new models for the spread of agriculture. In: *Proceedings-British Academy*, vol. 144, Oxford University Press Inc., p. 99.
- Rotunno, R., Crema, E.R., 2025. Bayesian analyses of radiocarbon dates suggest multiple origins of ceramic technology in early holocene africa. *Nat. Commun.* 16 (1), 8819.
- Rue, H., Held, L., 2005. *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC.
- Russell, T., Silva, F., Steele, J., 2014. Modelling the spread of farming in the bantu-speaking regions of africa: An archaeology-based phylogeography. *PLoS One* 9 (1), e87854. <http://dx.doi.org/10.1371/journal.pone.0087854>, Publisher: Public Library of Science. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0087854>.
- Shennan, S., 2001. Demography and cultural innovation: a model and its implications for the emergence of modern human culture. *Camb. Archaeol. J.* 11 (1), 5–16.
- Silva, F., Stevens, C.J., Weisskopf, A., Castillo, C., Qin, L., Bevan, A., Fuller, D.Q., 2015. Modelling the geographical origin of rice cultivation in Asia using the rice archaeological database. *PLoS One* 10 (9), e0137024.
- Team, R.D.C., 2020. R: A language and environment for statistical computing. R Found. Stat. Comput. Vienna, Austria URL <https://www.r-project.org/>.
- Thomas, M., 1949. A generalization of Poisson's binomial limit for use in ecology. *Biometrika* 36 (1/2), 18–25.
- Van Etten, J., Hijmans, R.J., 2010. A geospatial modelling approach integrating archaeobotany and genetics to trace the origin and dispersal of domesticated plants. *PLoS One* 5 (8), e12060.
- Vansina, J., 1990. *Paths in the Rainforests: Toward a History of Political Tradition in Equatorial Africa*. University of Wisconsin Press.
- Ward, G.K., Wilson, S.R., 1978. Procedures for comparing and combining radiocarbon age determinations: a critique. *Archaeometry* 20 (1), 19–31.
- Wheeler, D.C., Waller, L.A., 2008. Mountains, valleys, and rivers: the transmission of raccoon rabies over a heterogeneous landscape. *J. Agric. Biol. Environ. Stat.* 13 (4), 388–406.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., et al., 2016. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3 (1), 1–9.
- Womble, W.H., 1951. Differential systematics. *Science* 114 (2961), 315–322.