9

NON-STATIONARITY AND LOCAL SPATIAL ANALYSIS

Enrico R. Crema

Introduction

One of the core assumptions hold by most spatial analyses is that the generative process behind the observed pattern is *stationary*. This implies that statistical properties such as the intensity of a point process, the nature of the relationship between dependent and independent variables, or the patterns of spatial interaction are independent of their absolute location, and hence *homogenous* across space. The assumption is often adopted implicitly and not exclusively in spatial analyses; for example, when inferring population trajectories of a particular region (using site counts or density of radiocarbon dates), the pattern observed in the aggregate time-series is considered to be, at least to some extent, representative of the region as a whole. The advantage of holding this view is that information can be reduced into *global statistics*, enabling for example the description of complex and multi-scalar patterns of spatial interaction using a single, distance-based function (cf. Bevan this volume, Figure 4.3). Yet in many cases holding such an assumption might be problematic as many processes do vary in their properties across geographic space. They are, in other words, spatially *heterogeneous* and *non-stationary*. Under these circumstances choosing inappropriate methods that assume stationarity might at best hinder the detection of interesting variations and outliers in the data, and at worst lead to an erroneous understanding of the overall pattern.

The common way to informally approach potential issues derived from *non-stationarity* is to simply select a window of analysis where the generative process can be assumed to be spatially homogeneous. Intuitively speaking stationarity is negatively correlated with scale as larger study areas are more likely to incorporate variation in spatial properties, making the use of global statistics less appropriate. The problem is that the exact scale where the assumption stops being valid can vary depending on the nature of the process under investigation and the idiosyncrasies of the specific case study. While informal rules of thumb might be appropriate in some situations, stationarity should not be an *a priori* assumption, but rather a hypothesis to be evaluated. This is particularly the case for large scale synthetic research that harnesses the availability of increasingly larger collections of digital data, ranging from spatial databases of radiocarbon dates (e.g. Shennan et al., 2013; Chaput et al., 2015) to remotely sensed data (e.g. Menze, Ur, & Sherratt, 2006; Biagetti et al., 2017).

It is worth noting that *stationarity* is a property of the model, and not of the observed data per se (Fortin & Dale, 2005). This is a crucial point, as in practical terms *non-stationarity* arises from model

misspecification. Stationarity is an assumption whereby statistical properties such as mean or variance are considered to be spatially invariant. But these properties do often vary over space as a consequence of some unidentified variables, and failing to appropriately model these will drastically reduce our capacity to explain spatial heterogeneity and lead to the incorrect use of global statistics. (Fotheringham, Brunsdon, & Charlton, 2000). A trivial example can explain this issue. Suppose that someone is analysing the regional distribution of archaeological sites over a rugged landscape characterised by patches of flat areas that are more suited for human occupation. For the sake of simplicity, we can assume that the only driver of site density is the terrain morphology, with the intensity of occupation being five times larger in the flat patches. The density of archaeological sites will not be homogenous over space, but instead characterised by several clusters located in these patches. Examining this data and computing a single estimate of site density (i.e. computing a global statistic) would be inappropriate, and similarly analysing for spatial interaction might misleadingly suggest evidence of second-order interaction (when in fact the sites are not attracted to each other but only to absolute spatial locations). The problem can be solved by either analysing the flat patches separately by partitioning the study area or by *specifying* a variable that explains the variation in site density (i.e. terrain ruggedness). Ignoring either option will lead to incorrect inferences.

The substantial growth in the availability of Geographic Information Systems (GIS)-based spatial data in recent years has undoubtedly eased the creation of more sophisticated and complex models that can account for different kinds of spatial dependencies induced by environmental variables. If appropriately identified and modelled, these advances can limit the risk of model misspecifications. However, this is not necessarily a trivial task, and the situation is worsened for two reasons. First, spatial differences can also arise simply as a consequence of heterogeneity in archaeological research design. Different states, regions, and individuals often employ different sampling strategies resulting in biases that can exhibit strong spatial structure. Figure 9.1, for example, shows the location of North American archaeological sites included in the Canadian Archaeological Radiocarbon Database (CARD, v.2.0). The overall variation in the density of archaeological sites with radiocarbon dates is a combined effect of past population density and differences in sampling intensity, but the remarkable strength of the latter is not always as self-evident as in the case of the state of Wyoming, shown here as a rectangular patch with a disproportionately high sample density. Despite the known role of these forms of sampling biases, archaeological spatial analyses have rarely addressed this issue formally (but see Bevan (2012) for an exception; see also Banning, this volume). Yet examples in fields such as ecology showcase how the challenging task of quantifying and formally integrating sampling bias is not only possible but can dramatically improve the predictive power of a model (see Syfert, Smith, and Coomes (2013) and Stolar and Nielsen (2015) for applications in species distribution modelling).

Second, whilst model misspecification and sampling bias are, at least potentially, tractable problems, non-stationarity can also arise because different individuals might genuinely exhibit different relationships across space. Cultural, behavioural, and economic differences can in fact lead to different practices, attitudes, and preferences towards the very same environmental variable, and at the same time these variations are likely to exhibit spatial autocorrelation. Global analysis will, by definition, ignore these potential variations as its core assumption is that individual observations are interchangeable and originating from the same process. This can be regarded as a particular form of model misspecification (e.g. one could, at least in theory, specify categorical variables to depict cultural affiliations), albeit one where identifying and quantifying key variables is difficult, if not impossible. From a theoretical standpoint ignoring potential spatial heterogeneity arising from these factors is an example of environmental determinism (see Gaffney & van Leusen, 1995; Jones & Hanham, 1995), an approach that 'denies geography and history' by assuming that 'every time and everywhere is basically the same' (Jones 1991, p8; cited in Fotheringham et al., 2000, p. 95).

FIGURE 9.1 Screenshot depicting the distribution of radiocarbon dates available from the Canadian Archaeological Radiocarbon Database, version 2.1 (Martindale et al., 2016). A colour version of this figure can be found in the plates section.

Method

How then can we identify non-stationarity? How can we discern cases where using global statistics is still appropriate in contrast to instances where model misspecifications, sampling bias, and un-modelled cultural variables can deeply undermine the results of the spatial analysis? Within a typical modelling framework (e.g. regression analysis), the standard way to tackle this issue is to examine for the presence of spatial autocorrelation in model residuals. While this is an efficient solution that directly examines the assumptions of global statistics, the detection of spatial structure in the residual provides only a general indication of misspecification and does not provide sufficient details on the nature of the spatial variation *per se*.

One way to approach this problem is to break down the average properties observed at the global scale and focus the perspective on to its local scale constituents. Thus rather than yielding a single statistic describing the entire window of analysis, the objective is to retrieve multiple values, one for each of the sampled locations. By analysing these statistics or even simply visualising them on a map, regularities and exceptions can be identified. This provides clues for identifying plausible missing variables or provides some insights into the nature of a culturally-driven spatial heterogeneity. The growth of global information systems (GIS) in the mid-90s has particularly fostered the development of a suite of statistical

techniques, generally referred to as *local spatial analysis*, that implements this shift from a global to a local perspective. These include both local versions of pre-existing global statistics (e.g. Local Ripley's K, Local Moran's I, Geographically Weighted Regression, Spatial Expansion Method, etc.) as well as purposely developed new methods (e.g. the geographical analysis machine, GAM, by Openshaw, Charlton, Wymer, & Craft, 1987, but also the *locally-adaptive model of archaeological potential*, LAMAP, by Carleton, Conolly, & Iannone, 2012).

While these techniques vary in their details (see below), they generally share two main properties: (1) statistics are computed for each observed sample location, and hence they can be "mapped"; and (2) statistics are computed by weighting the contribution of samples based on the distance to each focal observation, i.e. they are based on *local* neighbourhoods that can be specified in a variety of ways (e.g. contiguity in polygon data, a fixed number of 'nearest' neighbours, cut-off distance, distance decay functions etc. . . . (see Getis & Aldstadt, 2010, for a review). The subsections below provide a brief summary of the key concepts pertaining to the most commonly used forms of local spatial analysis, and a review of their archaeological applications.

Local point pattern analysis

Point pattern analysis (see Bevan, this volume) refers to a body of statistical techniques designed to assess the spatial distribution of entities that can be described as points located most typically (but not necessarily) on a two-dimensional plane. The underlying processes behind a given point pattern are determined by exogenous and/or endogenous factors. The former is often referred to as a *first-order effect* (Bailey & Gatrell, 1995), or *induced spatial dependency* (Fortin & Dale, 2005) and includes factors that are independent from the phenomena of interest such as topography, soil, or distance to key resources. Endogenous factors are instead referred to as a *second-order effect* (Bailey & Gatrell, 1995), or *inherent spatial dependency* (Fortin & Dale, 2005). These include factors that are intrinsic to the phenomena of interest, such as the repulsion between settlements resulting from territoriality, or the aggregation of house-units driven by socio-economic principles. The goal of point pattern analysis is to discern and model these two forms of dependency.

Archaeological applications of point pattern analysis have a long tradition that goes back to the early 1970s (see references within Hodder & Orton, 1976), and since then the focus has been predominantly on methods designed to discern between regular, clustered, and random patterns such as the Nearest Neighbour Index (Clark & Evans, 1954) or the Ripley's K function (Ripley, 1976). Both of these methods are global statistics, and while they provide easy to interpret numerical indices and can be used within a hypothesis testing framework, they generally assume stationarity and do not directly distinguish induced and inherent spatial dependency. This can be problematic in a variety of ways. First, the standard null hypothesis used in most techniques is a spatially homogeneous Poisson process where the *intensity* (i.e. density) is estimated from the observed data. Techniques such as Ripley's K function or the Pair Correlation Function (PCF) are designed to detect spatial interaction (i.e. inherent spatial dependency) as instances of clustering or dispersion that are not accounted for by the null model (cf. Figure 9.2(a) vs. Figure 9.2(b); see Bevan, this volume). Both techniques extrapolate a measure of local density at different spatial scales and compare them against expectations from this null model. Statistical significance is then obtained by simulating point patterns under the null hypothesis and generating a simulation envelope: observed statistics above this envelope are then interpreted as evidence of clustering at the given spatial scale, whilst statistics below the envelope are regarded as evidence of dispersion (see Bevan, this volume for further details).

However, clustering can also be the result of induced spatial dependency that is often expected in a larger window of analysis (e.g. settlement clustering along rivers, or on flat patches – cf. the earlier

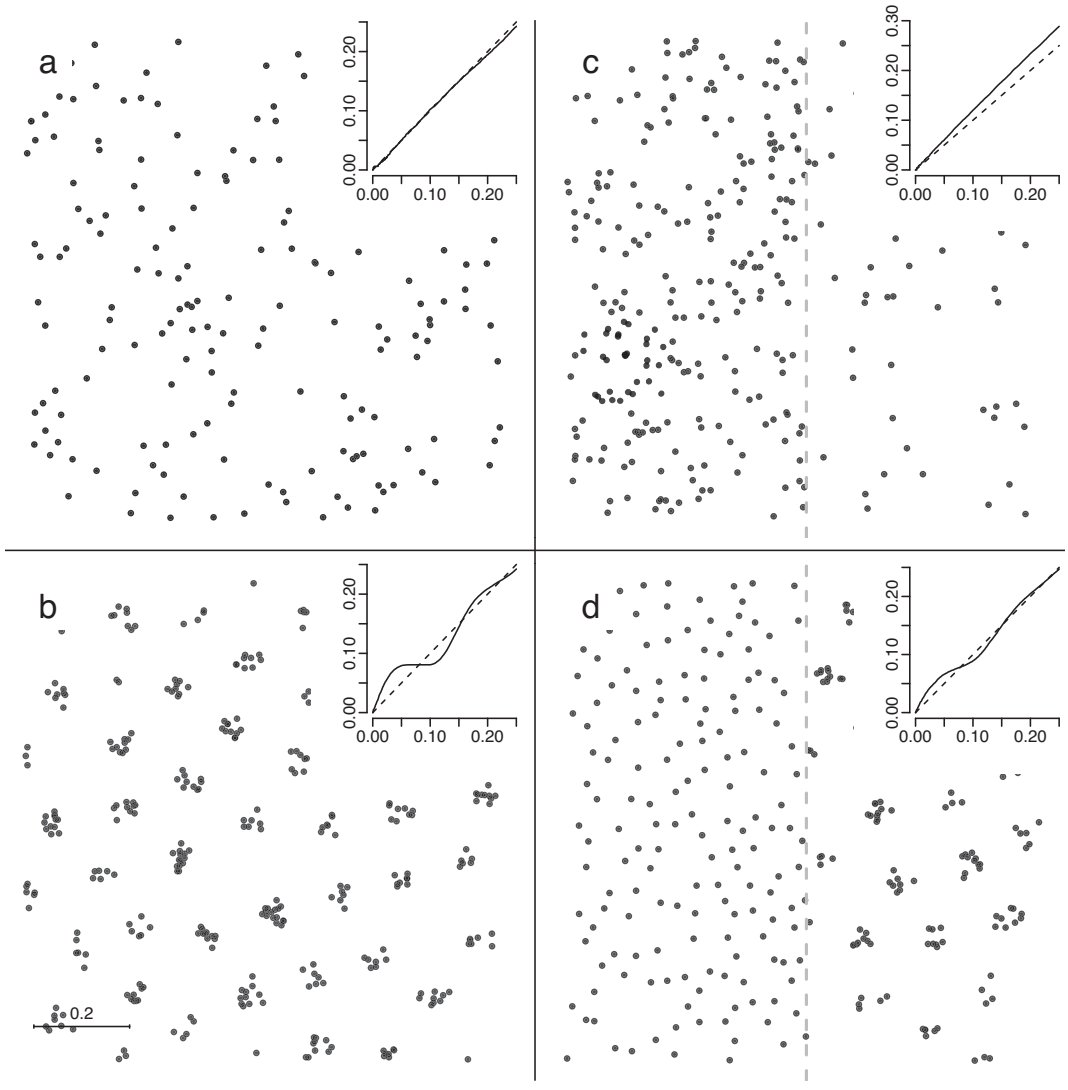

FIGURE 9.2 Simulated point patterns with associated observed (solid line) and expected (dashed line, under Complete Spatial Randomness) L function (a variant of Ripley's K function where the theoretical expectation of Complete Spatial Randomness (CSR) is a straight line): (a) homogeneous Poisson process; (b) clustered point process; (c) spatially inhomogeneous Poisson process with different intensities between left and right sides of the window of analysis (separated by the dashed line); (d) second-order spatial heterogeneity with a combination of regular (left) and clustered (right) patterns. The function suggests aggregation (clustering) when the observed L function is above the expected value and segregation (regular spacing) when below.

example). If the objective is the detection of spatial interaction then using a homogeneous Poisson process in this case can be regarded as a particular form of misspecification (see Figure 9.2(c)). The issue can be tackled by using more sophisticated techniques that can replace the null hypothesis with a spatially *inhomogeneous* version of the Poisson model, where the intensity varies as a function of external covariates.

For example, Eve and Crema (2014) investigated the distribution of Bronze Age houses at Leskernick Hill (Cornwall, UK) by first fitting a point process model using a range of covariates including elevation, slope, and visibility of landmarks (i.e. modelling induced spatial dependency), and subsequently used a residual K function to detect clustering that was not accounted for by their fitted model (i.e. inherent spatial dependency).

This solution is feasible as long as the *inhomogeneous* Poisson model can be assumed to be stationary. However, the *relationship* between the intensity of the point process and the external variables (described by the parameters of the fitted model) might also vary over space. If this is the case a global fitted model is no longer a viable option and one should adopt alternative solutions (see Baddeley (2017)) similar to those used in geographically weighted regression (see below).

Furthermore, even when variation in the externally induced spatial dependency is taken into account, the nature of spatial interaction (i.e. inherent spatial dependency) might still vary over space (Figure 9.2(d)). Such second-order heterogeneity (Pélissier & Goreaud, 2001) cannot be tackled by the most commonly adopted point-pattern analysis techniques such as Ripley's K function or Nearest Neighbour Index, as the mathematics underpinning the methods described above are based on aggregate statistics (e.g. the mean density within a specific radius or the average distance to the nearest neighbour) that effectively ignore variation between observations.

The solution in this case is to measure the same statistic for each observation point and map their variation over space. The most widely adopted example of this approach is Getis and Franklin's (1987) *second-order neighbourhood analysis*, which is effectively equivalent to a local version of Ripley's K function. A few archaeological examples employ this technique either in its basic form (e.g. Palmisano, 2013) or in its bivariate version, where the inherent spatial dependency is investigated in terms of relationship attraction or repulsion between two classes of points (e.g. two different artefact types). For example, Orton (2004) re-examined the flint artefact distribution within the Mesolithic site of Barmose I identifying potential activity areas as an alternative to cluster analyses. Crema and Bianchi (2013), and more recently Riris (2017), applied the same suite of techniques on survey data, operationalizing the transition from a site-centric to artefact-centric analysis of surface scatters. Both of these studies identified local patterns of inter-type artefact aggregation and segregation (with statistical significance obtained from random permutation tests), and more importantly 'mapped' the variation of such relationships over space, identifying complex patterns within and between clusters that cannot be adequately described by standard global spatial analysis. Figure 9.3 compares, for example, the output of a global (Figure 9.3(b)) and a local (Figure 9.3(c)) point pattern analysis aimed to assess the aggregation/segregation of stone tools made of different raw materials (see Crema & Bianchi (2013) for further details). The global bivariate L function suggest an aggregation between different materials (in this case Gafsa sourced flint vs flint sourced from elsewhere) up to 350 meters. The local version of the same analysis shows, however, that this aggregation occurs only in some areas (see filled dots in Figure 9.3(c)).

Local indicators of spatial association (LISA)

One of the most commonly adopted forms of local spatial analysis is a suite of geostatistical techniques commonly referred to as *local indicators of spatial association* (LISA; see also Fusco and de Runz, this volume). These are designed to determine for any given sampled location and its local neighbourhood the extent of clustering of similar observed values (Anselin, 1995). The primary objective of LISA is thus to decompose global indices of spatial autocorrelation into their local constituents in order to identify the location of outliers and local spots of non-stationarity. Although the various association statistics (i.e. local Gamma, local Moran, and local Geary), differ slightly from each other they generally all employ Monte-Carlo

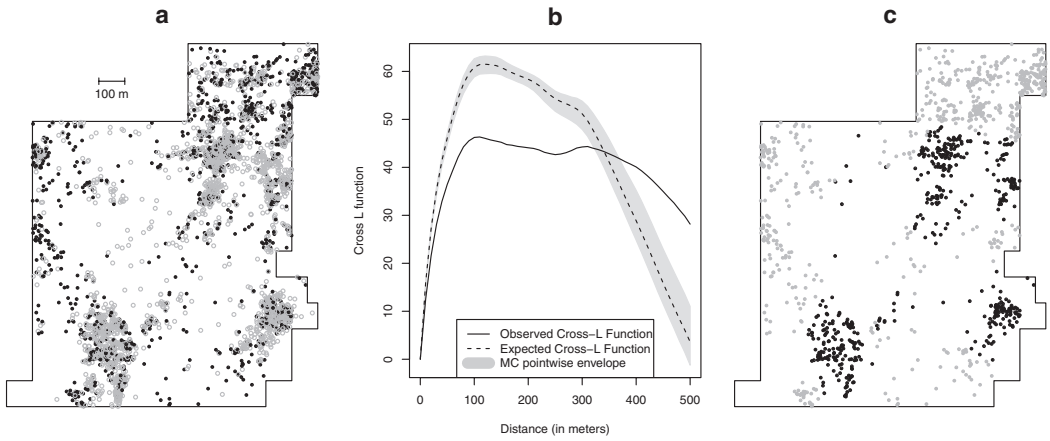

FIGURE 9.3 Lithic distribution analysis from the Sebkha Kelbia survey, Tunisia (after Crema & Bianchi, 2013), showing contrasting results between global and local bivariate L functions of stone tools divided by their raw material (Gafsa flint vs. flint sourced elsewhere): (a) distribution of the analysed stone tools (filled circle: Gafsa sourced flint; hollow circle: flint sourced from elsewhere); (b) bivariate L function showing significant segregation between the two classes between 20 and 320 meters (MC: Monte-Carlo); (c) local bivariate L function scale showing evidence of aggregation at 100 meters (black dots indicate location of Gafsa sourced flints with a statistically significant proportion of neighbours composed by flint sourced from elsewhere).

simulations to assess statistical significance and formally define spatial neighbourhoods using a weighted scheme (Getis & Aldstadt, 2010).

While the most conventional use of LISA is to provide a better diagnostic tool of regression analysis by identifying *where* residuals exhibit strong autocorrelation, the range of archaeological applications testifies how this suite of techniques (along with other local versions of geostatistical analyses) can be used in a range of contexts. For example, Premo (2004) used Moran's local *I* (Anselin, 1995) and Getis's local G_i* statistics (Getis & Ord, 1992; a related technique designed to identify clusters distinguishing whether they are low or high values compared to the mean) to explore the spatial distribution of terminal long-count dates carved on Classic Maya monuments. The objective in this case was to determine whether these proxies of 'collapse' (the terminal dates indicate the most recent year when elites at a particular site raised monuments) exhibit local variations in their extent of autocorrelation, and identify the presence and the location of significant clusters of early and late dates. Crema, Bevan, and Lake (2010) also used the local G_i* statistics as an exploratory analysis to identify areas of low or high chronological uncertainty in Middle to Late Jomon pit-dwellings in central Japan. More recently Styring, Maier, Stephan, Schlichtherle, and Bogaard (2016) used the same analysis on the $\delta^{15}N$ value of cereal grains at the Neolithic site of Hornstaad-Hörnle IA, Germany, to investigate patterns of inter-household variation in crop-husbandry practices.

Geographically weighted regression

Regression analyses are one of the most widely used statistical techniques in archaeology and often entail observations that are spatially situated (Hacıgüzeller, this volume). Typical examples include estimates of the speed of the spread of farming using radiocarbon dates (e.g. Pinhasi, Fort, & Ammerman, 2005),

the fitting of fall-off curves of proportion data (e.g. artefact type) from potential centres of production (e.g. Eerkens, Spurling, & Gras, 2008), or the modelling of site presence/absence via logistic regression (e.g. Carrer, 2013; see Kvamme this volume). Most of these regression models assume that: (a) samples are independent; and (b) the observed relationships between variables are the same across space, i.e. they assume stationarity. The latter implies that estimates of the rate of expansion are assumed to be constant over space, decrease in the proportion of artefact types from the source is assumed to be *isotropic* (i.e. there is no directionality in the fall-off), and that the independent variables are assumed to have the same role in determining the likelihood of site presence across the study area. As for the other cases, if these assumptions are not justified, models can potentially be misspecified and estimates biased.

While diagnosis of regression residuals can help identify problematic cases, they do not explicitly model spatial heterogeneity and hence do not provide means to formally approach the non-stationarity problem (i.e. they are not able to inform *how* the relationships vary across space). The last two decades however have seen the development of a wide range of regression techniques designed for the analyses of spatial data. Problems such as the non-independence and autocorrelation of sample observations are being tackled by tailored methods such as spatial auto-regressive models (see Gil et al., 2016 for an archaeological application). *Geographically Weighted Regression* (GWR) (Fotheringham, Brunsdon, & Charlton, 1998; Fotheringham et al., 2002) is one such technique that is suited for instances where the relationship between variables are known to be spatially heterogeneous. The method is essentially a 'local' version of regression analysis where global model parameters are replaced by continuous functions that are dependent on the spatial coordinates of each location. Thus a 'global' regression model can be regarded as a special case of the GWR, whereby the output of these continuous functions do not vary across space. By allowing model parameters to vary across space this technique takes into account spatial heterogeneity (reducing model misspecification), and allowing at the same time the possibility to 'map' the spatial variation of the parameters (and hence the spatial variation in the relationship between dependent and independent variables). Geographically weighted regression assumes that when estimating the parameters for a given location *i*, sites in proximity have a larger impact in the estimate of the model parameters than those that are further away. This is achieved by weighting the contribution of neighbouring data points using some distance decay function. Geographically weighted regression shares some similarities with the *spatial expansion method* (Jones & Casetti, 1992), an earlier technique that similarly highlighted the importance of spatially varying relationships. Whilst the *spatial expansion method* is a relevant precursor of GWR, it provides less flexibility in defining *how* parameters vary over space, as it is designed to capture general directional trends and its form needs to be assumed a priori (Fotheringham et al., 2000).

Despite its ability to address potential issues of environmental determinism, archaeological applications of GWR have been comparatively limited. Gkiasta, Russell, Shennan, and Steele (2003) explored local variations in the rate of the spread of farming in Neolithic Europe, whilst Bevan and Conolly (2009) examined how covariates such as slope, vegetation, and geology have different relationships to the surface pot-sherd density in different parts of the Greek island of Antikythera using a geographically weighted zero-inflated Poisson regression. The technique has also been explored in the context of predictive modelling of site locations (Löwenborg, 2010), as well as for larger synthetic research such as Linearbandkeramik (LBK) faunal remains in western Europe (Manning et al., 2013).

Case study

The core principles shared across the methods described above can be applied to virtually any analysis that seeks to tackle non-stationarity. One recent archaeological example is the spatial extension of the summed probability distribution of radiocarbon dates (SPDRD). The non-spatial version of this technique has

recently renewed a strong interest in prehistoric demography, as the increasing availability of a large collection of radiocarbon dates is providing a new proxy for inferring past population trajectories within an absolute chronological framework. While the core assumptions of this "dates as data" (Rick, 1987) approach are still being discussed, it is undeniable that SPDRD is quickly becoming part of the standard toolkit in regional studies. In particular, the production of demographic time-series within an absolute chronology is opening new possibilities to infer the role of past climatic change (e.g. Kelly, Surovell, Shuman, & Smith, 2013; Warden et al., 2017) or to explore cross-regional divergences in demographic trajectories (e.g. Timpson et al., 2014; Crema, Habu, Kobayashi, & Madella, 2016), potentially at the global level (Chaput & Gajewski, 2016).

The possibility to incorporate a spatial dimension is particularly noteworthy here as it requires a careful balance between sample size and the spatial extent of the window of analysis. Because the shape of SPDRD is subject to sampling error, a formal assessment of its shape (i.e. the hypothesised demographic trajectories) will require a sufficient number of radiocarbon dates. While some suggestions for a threshold size have been suggested (e.g. Williams, 2012), the optimal size ultimately depends on the specific null hypothesis that is being tested (the most common ones being exponential and logistic population growths) and the effect size being sought. With other things being equal, the most straightforward solution to increase the sample size is to expand the size of the window of analysis. This, however, means that stationarity is harder to justify as different regions are likely to experience heterogeneous demographic histories (cf. sub-regions in Shennan et al., 2013 and Timpson et al., 2014) as well as different sampling strategies (see Figure 9.1, Bevan et al., 2017; see also Banning, this volume). The latter in particular hinders the straightforward application of methods such as Kernel Density Estimates (KDE; see Bevan, this volume), as the number of radiocarbon dates is determined at least in part by local differences in sampling intensity. Attempts to overcome this issue have been rare, with the notable exception of Chaput and Gajewski (2016) who employ relative risk surfaces (see also supplementary materials in Bevan et al., 2017) by taking the ratio of each KDE map by the overall sampling intensity. While this approach is a valuable correction in the observed pattern it does not distinguish between genuine instances of spatial heterogeneity from variations arising from sample error.

Crema, Bevan, and Shennan (2017) have recently explored this issue by developing a local spatial analysis designed to identify presence of spatial heterogeneity in the demographic trajectories hypothesised from the SPDRDs, enabling the formal assessment of non-stationarity. The method involves the following six steps (for the full description see the original paper):

1) Compute for each site *i* a local SPDRD which is created by summing all radiocarbon probabilities but weighting (using an exponential decay function) the contribution of dates from neighbouring sites as function of distance from *i*.
2) Define temporal slices (e.g. 7500–7001 cal BP, 7000–6501 cal BP, etc.) and compute the geometric growth rate between abutting pairs for each local SPDRD (e.g. between 7500–7001 and 7000–6501 cal BP, between 7000–6501 cal BP and 6500–6001 cal BP, and so on . . .).
3) Randomly permute the spatial coordinates of the radiocarbon dates, so that the entire set of dates associated to a particular location *x* is given a new location *y*, and then execute steps 1 and 2 above.
4) Repeat step 3 *n* times, so that for each transition (e.g. from 7500–7001 to 7000–6501 cal BP) at each site, there is one *observed* geometric growth rate (obtained in steps 1–2), and *n simulated* geometric growth rates (obtained in step 3). The latter is the expected pattern under the assumption of spatial stationarity (i.e. the same expected growth rate across space with variation entirely determined by sampling error). Notice that the envelope of the simulated dates will be narrower in regions with a higher sampling intensity and wider in areas with a lower sampling intensity.

5) Compare the observed and simulated growth rates for each location and compute the p-value for significance testing, equivalent to $(r+1)/(n+1)$ where r is the number of replicates where the simulated growth rate is lower (or higher) than the observed rate.
6) Use the distribution of p-values to compute *false discovery rates* (q-values, Benjamini & Hochberg, 1997) to take into account expected inflation of type I error (i.e. incorrect rejection of a true null hypothesis) due to multiple hypothesis testing.

Figure 9.4 shows the result of this local analysis applied in the context of Neolithic Europe. The red dots indicate site locations with a significant (q-value < 0.05) local positive departure from the expected growth rate under stationarity in the transition between 6500–6001 cal BP and 6000–5500 cal BP (transition IV), whilst the blue dots indicate the opposite (lower than expected rate). If all regions experienced similar population trajectories (as inferred from the density of radiocarbon dates) and local variations in the SPD were purely the result of sampling error, we would not expect to observe any significant positive or negative departures. The insets show the result of two particular locations where the observed growth rate (solid line with filled dots) is higher and lower than the expected rates under stationarity (dashed line

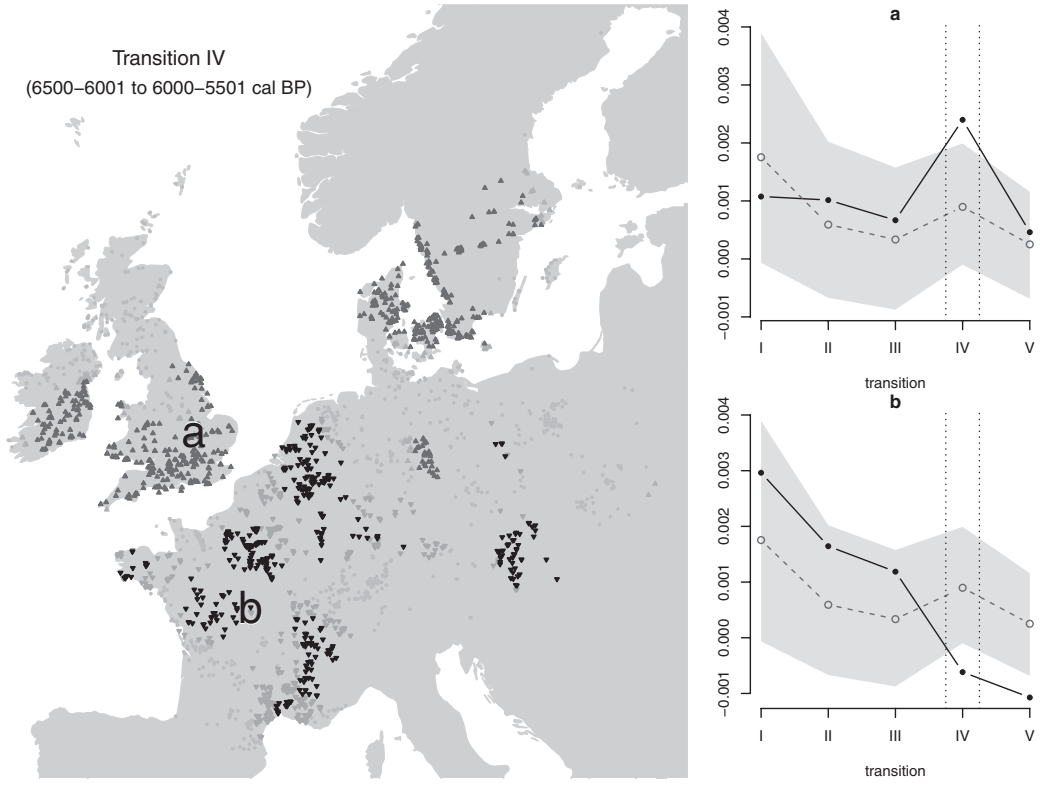

FIGURE 9.4 Local spatial permutation test of the summed probability distribution of radiocarbon dates (SPDRD) from Neolithic Europe showing locations with higher (red) or lower (blue) geometric growth rates than the expectation from the null hypothesis (i.e. spatial homogeneity in growth trajectories) at the transition period between 6500–6001 and 6000–5501 cal BP. The insets on the right show the observed local geometric growth rates and the simulation envelope for locations a and b on the map (see Crema et al., 2017, for details). A colour version of this figure can be found in the plates section.

with hollow dots) and its associated simulation envelope (grey region) obtained from 10,000 permutations. The result indicates statistically significant instances of spatial heterogeneity in growth rates, with southern Britain, southern Ireland, the Baltic regions, and parts of central Germany experiencing higher growth rates while most of continental Europe within the study area have the opposite pattern.

Conclusion

The substantial heterogeneity in the objectives, the type of data, and the scale of analysis, makes the application of spatial analysis in archaeology a challenging and diverse task. Techniques are mostly developed in other fields and come with assumptions that were valid for the particular contexts they were designed for. Whilst generalised tools are highly desired, the underpinning assumptions are not easily transferrable across different applications. The problem is exacerbated by the fact that too often we ignore the assumptions and their implications entirely, leading to a divergence between archaeological theories and spatial models.

The problem of non-stationarity is a good example of this; the majority of spatial statistics used in archaeology assume spatial homogeneity, yet the theoretical stance and interest of archaeologists is often focused much more on heterogeneity. Despite the availability of a substantial range of techniques that are designed to tackle non-stationarity (or to model spatially heterogeneous processes), archaeological applications are comparatively rare with global statistics still being the most commonly adopted approach. The growing amount of high quality data at increasingly larger spatial scales might however change this and promote the use of local spatial analysis. This will no doubt provide new perspectives on the human past, enabling us to answer questions that are perhaps in line with a wider range of theoretical approaches. Such a shift in scale will, however, require the creation of more tailored techniques as well as the retrieval of data that can provide the basis for exploring the effects of research bias. It is undeniable that with the increasing possibility to engage with larger spatial scales, we will have to face the impact of heterogeneous research practices. These will have a greater role in shaping the distributions we observe, hindering our ability to isolate the patterns we truly seek to study. The adoption of local statistics can help this endeavour but it is worth noting that these are ultimately exploratory tools and can never replace a global model where key missing variables are correctly integrated. Detecting spatial heterogeneity tells us only that there is something missing; we might estimate *where* and to some extent even *how*, but it will never tell us *what* they are. Furthermore, one should also avoid the temptation to exclusively rely on the inductive insights offered by the output of local analyses and conceiving them as the final stage of a research workflow. This is particularly so because the number of statistical hypotheses is generally as many as the number of observations. As a consequence, there is an increased possibility of incorrectly rejecting the null hypothesis even when this is false (type I-error). This is a known problem and one that cannot be easily solved by standard correction methods, such as Bonferroni, as tests are not entirely independent from each other and consequently an indiscriminate use of p-value adjustment can lead to overly conservative conclusions (i.e. type II errors). This is also a known issue within the literature of local spatial analysis (e.g. de Castro & Singer, 2006), and while some suggestions have been proposed there is no consensus towards a single solution. Ultimately, local analyses should not be considered as substitutes for global statistics but rather as a suite of complementary tools for evaluating assumptions, providing clues for searching for missing variables, and refining hypotheses.

Acknowledgements

I would like to thank Mark Gillings, Gary Lock, and Piraye Hacıgüzeller for inviting me, providing constructive feedbacks to the manuscript, and above all for being patient. I am also grateful to endless

discussions on these topics with a number of colleagues, in particular Andrew Bevan, Mark Lake, and Alessio Palmisano. Analyses were performed using the *spatstat* (Baddeley, Rubak, & Turner, 2015) and *rcarbon* (Bevan & Crema, 2018) packages within R statistical computing language (R Core Team, 2018).

References

Anselin, L. (1995). Local indicators of spatial association–LISA. *Geographical Analysis*, 27, 93–115.

Baddeley, A. (2017). Local composite likelihood for spatial point processes. *Spatial Statistics*, 22, 261–295.

Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial point patterns: Methodology and applications with R*. London: Chapman and Hall/CRC Press.

Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis*. Harlow: Prentice Hall.

Biagetti, S., Merlo, S., Adam, E., Lobo, A., Conesa, F. C., Knight, J., . . . Madella, M. (2017). High and medium resolution satellite imagery to evaluate late Holocene human–environment interactions in Arid lands: A case study from the Central Sahara. *Remote Sensing*, 9, 351.

Benjamini, Y., & Hochberg, Y. (1997). Multiple hypotheses testing with weights. *Scandinavian Journal of Statistics*, 24m, 407–418.

Bevan, A. (2012). Spatial methods for analysing large-scale artefact inventories. *Antiquity*, 86, 492–506.

Bevan, A., Colledge, S., Fuller, D., Fyfe, R., Shennan, S., & Stevens, C. (2017). Holocene fluctuations in human population demonstrate repeated links to food production and climate. *Proceedings of the National Academy of Sciences of the United States of America*, 114(49), E10524–E10531.

Bevan, A., & Conolly, J. (2009). Modelling spatial heterogeneity and nonstationarity in artifact-rich landscapes. *Journal of Archaeological Science*, 36, 956–964.

Bevan, A., & Crema, E. R. (2018). *Rcarbon v1.2.0: Methods for calibrating and analysing radiocarbon dates*. Retrieved from https://CRAN.R-project.org/package=rcarbon

Carrer, F. (2013). An ethnoarchaeological inductive model for predicting archaeological site location: A case-study of pastoral settlement patterns in the Val di Fiemme and Val di Sole (Trentino, Italian Alps). *Journal of Anthropological Archaeology*, 32, 54–62.

Carleton, W. C., Conolly, J., & Iannone, G. (2012). A locally-adaptive model of archaeological potential (LAMAP). *Journal of Archaeological Science*, 39, 3371–3385.

Chaput, M. A., & Gajewski, K. (2016). Radiocarbon dates as estimates of ancient human population size. *Anthropocene*, 15, 3–12.

Chaput, M. A., Kriesche, B., Betts, M., Martindale, A., Kulik, R., Schmidt, V., & Gajewski, K. (2015). Spatiotemporal distribution of Holocene populations in North America. *Proceedings of the National Academy of Sciences of the United States of America*, 112(39), 12127–12132.

Clark, P. J., & Evans, F. C. (1954). Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology*, 35, 445–453.

Crema, E. R., Bevan, A., & Lake, M. (2010). A probabilistic framework for assessing spatio-temporal point patterns in the archaeological record. *Journal of Archaeological Science*, 37, 1118–1130.

Crema, E. R., Bevan, A., & Shennan, S. (2017). Spatio-temporal approaches to archaeological radiocarbon dates. *Journal of Archaeological Science*, 87, 1–9.

Crema, E. R., & Bianchi, E. (2013). Looking for patterns in the noise: non-site spatial analysis at Sebkha Kalbia Tunisia. In S. Mulazzani (Ed.), *Le Capsien de hergla Tunisie Culture, environnement et économie* (pp. 385–395). Frankfurt: Africa Magna.

Crema, E. R., Habu, J., Kobayashi, K., & Madella, M. (2016). Summed probability distribution of 14 C dates suggests regional divergences in the population dynamics of the Jomon Period in Eastern Japan. *PLoS One*, 11. doi:10.1371/journal.pone.0154809

de Castro, M. C., & Singer, B. H. (2006). Controlling the False Discovery Rate: A new application to account for multiple and dependent tests in local statistics of spatial association. *Geographical Analysis*, 38, 180–208.

Eerkens, J. W., Spurling, A. M., & Gras, M. A. (2008). Measuring prehistoric mobility strategies based on obsidian geochemical and technological signatures in the Owens Valley, California. *Journal of Archaeological Science*, 35, 668–680.

Eve, S., & Crema, E. R. (2014). A house with a view? Multi-model inference, visibility fields, and point process analysis of a Bronze Age settlement on Leskernick Hill (Cornwall, UK). *Journal of Archaeological Science*, 43, 267–277.

Fortin, M.-J., & Dale, M. (2005). *Spatial analysis: A guide for ecologists*. Cambridge: Cambridge University Press.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (1998). Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 30, 1905–1927.

Fotheringham, S. A., Brunsdon, C., & Charlton, M. (2000). *Quantitative geography: Perspectives on spatial data analysis*. London: Sage Publications.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester: John Wiley & Sons.

Gaffney, C. F., & van Leusen, P. M. (1995). Postscript–GIS, environmental determinism and archaeology. In G. Lock & Z. Stančić (Eds.), *Archaeology and geographic information systems* (pp. 367–382). London: Taylor and Francis.

Getis, A., & Aldstadt, J. (2010). Constructing the spatial weights matrix using a local statistic. *Geographical Analysis*, 36, 90–104.

Getis, A., & Franklin, J. (1987). Second-order neighborhood analysis of mapped point patterns. *Ecology*, 68, 473–477.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24, 189–206.

Gil, A. F., Ugan, A., Otaola, C., Neme, G., Giardina, M., & Menéndez, L. (2016). Variation in camelid δ13C and δ15N values in relation to geography and climate: Holocene patterns and archaeological implications in central western Argentina. *Journal of Archaeological Science*, 66, 7–20.

Gkiasta, M., Russell, T., Shennan, S., & Steele, J. (2003). Neolithic transition in Europe: The radiocarbon record revisited. *Antiquity*, 77, 45–62.

Hodder, I., & Orton, C. (1976). *Spatial analysis in archaeology*. Cambridge: Cambridge University Press.

Jones, J. P., & Casetti, E. (1992). *Applications of the expansion method*. London: Routledge.

Jones, J. P., & Hanham, R. Q. (1995). Contingency, realism, and the expansion method. *Geographical Analysis*, 27, 185–207.

Kelly, R. L., Surovell, T. A., Shuman, B. N., & Smith, G. M. (2013). A continuous climatic impact on Holocene human population in the Rocky Mountains. *Proceedings of the National Academy of Sciences of the United States of America*, 110(2), 443–447.

Löwenborg, D. (2010). Using geographically weighted regression to predict site representativity. In B. Frischer, J. W. Crawford, & D. Koller (Eds.), *Making history interactive, proceedings of CAA conference, 37th annual meeting* (pp. 203–215), Williamsburg, VA and Oxford: Archaeopress.

Manning, K., Stopp, B., Colledge, S., Downey, S., Conolly, J., Dobney, K., & Shennan, S. (2013). Animal exploitation in the early Neolithic of the Balkans and central Europe. In S. Colledge, J. Conolly, K. Dobney, K. Manning, & S. Shennan (Eds.), *The origins and spread of domestic animals in southwest Asia and Europe* (pp. 237–252). Walnut Creek, CA: Left Coast.

Martindale, A., Morlan, R., Betts, M., Blake, M., Gajewski, K., Chaput, M., . . . Vermeersch, P. (2016). *Canadian Archaeological Radiocarbon Database (CARD 2.1)*. Retrieved April 1, 2017, from www.canadianarchaeology.ca/

Menze, B. H., Ur, J. A., & Sherratt, A. G. (2006). Detection of ancient settlement mounds. *Photogrammetric Engineering & Remote Sensing*, 72, 321–327.

Openshaw, S., Charlton, M., Wymer, C., & Craft, A. (1987). A Mark 1 geographical analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems*, 1, 335–358.

Orton, C. (2004). Point pattern analysis revisited. *Archeologia e Calcolatori*, 15, 299–315.

Palmisano, A. (2013). Zooming patterns among the scales: A statistics technique to detect spatial patterns among settlements. In G. Earl, T. Sly, A. P. Chrysanthi, P. Murrieta-Flores, C. Papadopoulos, I. Romanowska, & D. Wheatley (Eds.), *CAA 2012: Proceedings of the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA)* (pp. 348–356). Amsterdam: Amsterdam University Press.

Pélissier, R., & Goreaud, F. (2001). A practical approach to the study of spatial structure in simple cases of heterogeneous vegetation. *Journal of Vegetation Science*, 12, 99–108.

Pinhasi, R., Fort, J., & Ammerman, A. J. (2005). Tracing the origin and spread of agriculture in Europe. *PLoS Biology*. doi:10.1371/journal.pbio.0030410

Premo, L. (2004). Local spatial autocorrelation statistics quantify multi-scale patterns in distributional data: An example from the Maya Lowlands. *Journal of Archaeological Science*, 31, 855–866.

R Core Team. (2018). *R: A language and environment for statistical computing: R Foundation for Statistical Computing*. Vienna, Austria. Retrieved from: www.R-project.org/

Rick, J. W. (1987). Dates as data : An examination of the Peruvian Preceramic Radiocarbon Record. *American Antiquity*, 52, 55–73.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13, 255–266.

Riris, P. (2017). Towards an artefact's-eye view: Non-site analysis of discard patterns and lithic technology in Neotropical settings with a case from Misiones province, Argentina. *Journal of Archaeological Science: Reports*, 11, 626–638.

Shennan, S., Downey, S. S., Timpson, A., Edinborough, K., Colledge, S., Kerig, T., . . . Thomas, M. G. (2013). Regional population collapse followed initial agriculture booms in mid-Holocene Europe. *Nature Communications*, 4. doi:10.1038/ncomms3486

Stolar, J., & Nielsen, S. E. (2015). Accounting for spatially biased sampling effort in presence-only species distribution modelling. *Diversity and Distributions*, 21, 595–608.

Styring, A., Maier, U., Stephan, E., Schlichtherle, H., & Bogaard, A. (2016). Cultivation of choice: New insights into farming practices at Neolithic lakeshore sites. *Antiquity*, 90, 95–110.

Syfert, M. M., Smith, M. J., & Coomes, D. A. (2013). The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PLoS One*. doi:10.1371/journal.pone.0055158

Timpson, A., Colledge, S., Crema, E., Edinborough, K., Kerig, T., Manning, K., . . . Shennan, S. (2014). Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: A new case-study using an improved method. *Journal of Archaeological Science*, 52, 549–557.

Warden, L., Moros, M., Neumann, T., Shennan, S., Timpson, A., Manning, K., . . . Damsté, J. S. S. (2017). Climate induced human demographic and cultural change in northern Europe during the mid-Holocene. *Scientific Reports*, 7, 15251.

Williams, A. N. (2012). The use of summed radiocarbon probability distributions in archaeology: A review of methods. *Journal of Archaeological Science*, 39, 578–589.