



Research papers

The application of Local Indicators for Categorical Data (LICD) to explore spatial dependence in archaeological spaces

Francesco Carrer^{a,*}, Tomasz M. Kossowski^b, Justyna Wilk^b, Michał B. Pietrzak^c, Roger S. Bivand^d

^a School of History, Classics and Archaeology, Newcastle University, Armstrong Building, NE1 7RU Newcastle upon Tyne, UK

^b Faculty of Human Geography and Planning, Adam Mickiewicz University, Krygowskiego 10, 61-680 Poznań, Poland

^c Department of Econometrics and Statistics, Faculty of Economic Sciences and Management, Nicolaus Copernicus University in Toruń (NCU), Gagarina 13a, 87-100 Toruń, Poland

^d Department of Economics, Norwegian School of Economics, Helleveien 30, N-5045 Bergen, Norway

ARTICLE INFO

Keywords:

Spatial analysis

Local autocorrelation

LICD

Join-count statistics

Regular and irregular lattice

ABSTRACT

Global and local analyses of spatial autocorrelation are commonplace in spatial archaeology. However, they are exclusively focused on continuous numerical parameters, even though logical (presence/absence) and categorical parameters are equally frequent in archaeological research. Global tests of spatial dependence for categorical data are routinely used in other fields, and local versions of these tests (known as LICD) have recently been developed. This paper provides a detailed description of such methods, and presents the first adaptation and application of LICD to archaeological data. Different LICD versions have been tested on two case-studies: (a) an archaeological grid, with object presence or absence recorded for each cell; (b) Historic Landscape Characterisation, with the origin of character types (mediaeval, post-mediaeval or modern) recorded for each region. These examples seek to showcase and promote the use of LICD in landscape and intra-site archaeology.

1. Introduction

Tobler's First Law of Geography states that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). This fundamental law underlies the concept of spatial dependence, which was formally described as a relationship between one point and other points in space (Anselin, 1988, p. 13). Spatial dependence refers to spatial autocorrelation, a relationship between a variable and the geographical proximity of the locations where it is observed (Getis, 2008). Various methods for measuring spatial autocorrelation (spatial dependence) have been developed for nominal, ordinal and interval spatial data. The first global spatial autocorrelation measure, coefficient I , was proposed by Moran (1948, 1950). In the same period, Geary (1954) introduced the so-called Geary's ratio c . Influential books written by Cliff and Ord (1973, 1981) analysed the statistical properties of Moran's I and Geary's c . Besides, Cliff and Ord delivered a precise description of *join-count* statistics for nominal spatial data (usually represented as k -coloured maps). Moran's I , Geary's c and *join-count* statistics represent different versions of the general Mantel's T statistic (Mantel, 1967). Unlike previous

works, Getis and Ord (1992) used distance statistics for measuring spatial autocorrelation. However, the application of Getis-Ord measures (G and G^*) has some limitations related to the maximum distance selection and the representation of distances in the form of a binary matrix. More recent developments in the analysis of spatial dependence include the entropy-based SG test, suitable for continuous and discrete data (López et al., 2010), and indices based on the conditional probabilities of surface object categories (Bai et al., 2016). A completely new measure of spatial association, based on a colocation quotient, has recently been proposed by Leslie and Kronenfeld (2011).

In the 1990s many scholars raised the question of how to measure the relationship between single locations and their neighbours. This led to the production of local spatial association measures, calculated for each location. Several papers published at that time extended the results from global spatial autocorrelation to the local case. Getis and Ord (1992) and Ord and Getis (1995) proposed local versions G_i , G_i^* of their global statistics. Anselin (1995) introduced local Moran's I_i and local Geary's c_i statistics, which are widely used as Local Indicators of Spatial Association (LISA). Recent developments in spatial

* Corresponding author.

E-mail addresses: francesco.carrer@newcastle.ac.uk (F. Carrer), tkoss@amu.edu.pl (T.M. Kossowski), jwilk@amu.edu.pl (J. Wilk), pietrzak@umk.pl (M.B. Pietrzak), Roger.Bivand@nhh.no (R.S. Bivand).

<https://doi.org/10.1016/j.jas.2020.105306>

Received 16 July 2020; Received in revised form 20 November 2020; Accepted 2 December 2020

Available online 15 December 2020

0305-4403/© 2020 Elsevier Ltd. All rights reserved.

statistics have provided new tools for addressing local spatial association for non-numerical (categorical) datasets. Originally introduced by Boots (2003, 2006), the methods known in the specialist literature as Local Indicators for Categorical Data (LICD) have been further developed by different research groups using distinct and complementary mathematical approaches (Bivand et al., 2017; Anselin and Li, 2019).

Assessing a degree of spatial association is particularly critical in archaeological research, as it might reveal complex social, cultural or depositional processes underlying spatial data patterning. Autocorrelation analysis has been commonplace in archaeological research since the early days of GIS (Kvamme, 1990), and different significance tests are now routinely applied in various archaeological projects at intra-site and inter-site (or landscape) scale (Carrer, 2017; Vaz, 2020).

Local methods are also well established in archaeology (Premo, 2004). Nevertheless, the main limitation of autocorrelation tests currently used in archaeology is that they all require the investigated parameter to be numerical and continuous. This prevents categorical variables from being tested for spatial dependence. The use of bivariate functions to assess the spatial interaction between marked point patterns (cross- K or cross- L) represents a valid and increasingly popular approach (Brandolini and Carrer, 2020; Biagetti et al., 2016), but it does not efficiently apply to point patterns where all potential locations are available, and it does not easily adapt to areal data with entities of variable size. *Join-count* statistics (Cliff and Ord, 1973) are widely used in different fields for this specific purpose (Epperson and Alvarez-Buylla, 1997; Hansen et al., 1997; Meng et al., 2005; Franklin et al., 2009), but to the best of our knowledge, they have almost never been applied to archaeological data. This might also be related to the fact that no effective local version of these tests were implemented until fairly recently, thus preventing its effective applicability to very heterogeneous datasets (like most archaeological datasets).

In this paper the potential of LICD for the investigation of patterning in archaeological spaces is explored for the first time. Two different approaches to LICD have been tested using the programming language R (R Core Team, 2020). In order to assess how these different approaches work at different scales and with different types of spatial data, two archaeological datasets have been chosen: (a) Historic Landscape Characterisation, GIS-based classification of the historical landscape character; (b) an excavation grid with a presence/absence record of artefact types per square metre.

2. Join-count statistics: Measuring of global autocorrelation for categorical data

In order to understand LICD methods and applications, an introduction to *join-count* statistics (JCS) is paramount. JCS is a simple yet solid method to measure the correlation between binomial variables and the distance between observations (Cliff and Ord, 1973). Let us consider a random variable X with $k \geq 2$ possible values (categories), which represent k events, and n being a number of spatial units. A realisation of X on the set of n spatial units can be presented on a k -coloured map showing the occurrence one of the k events in each spatial unit. In the simplest case $k = 2$, we have a two-coloured map (Fig. 1), where event regions are classified as black regions (B) and non-event regions as white (W). If the colour of a spatial unit i is black (B), then $x_i = 1$, otherwise $x_i = 0$. In order to test the positive autocorrelation of event (BB) and non-event cells (WW) (Fig. 1a), or the negative autocorrelation of events (BW) (Fig. 1c), the following equations are used:

$$\begin{aligned} BB &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j, \\ WW &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} (1 - x_i)(1 - x_j), \\ BW &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - x_j), \end{aligned}$$

where: w_{ij} is an element of the spatial weights matrix \mathbf{W} . It is assumed that each x_i is a realisation of Bernoulli distributed random variable

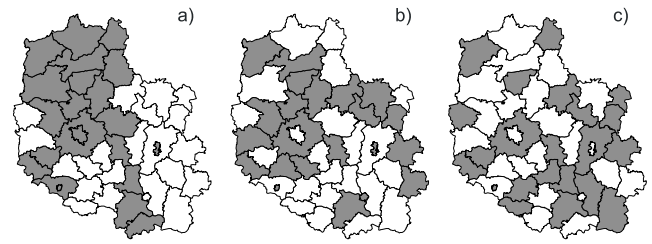


Fig. 1. Patterns on a two-coloured map: (a) positive spatial autocorrelation (clustering), (b) random distribution, (c) negative spatial autocorrelation (chessboard). Source: Kossowski (2018).

$X \sim B(n, p)$ where n is the number of spatial units, and p is the probability that, for given location i , $x_i = 1$ (Cliff and Ord, 1973; Pietrzak et al., 2014, p. 206). Since BB and BW statistics are asymptotically normally distributed, z -test of significance (Z_{BB} and Z_{BW}) can be used. In the case $k > 2$ we calculate three groups of statistics: (1) the same colour joins, (2) two different colours joins, (3) all different colours joins.

There are three strategies for testing the null hypothesis of random distribution of colours on the map against spatial autocorrelation among spatial units. The first approach assumes free sampling, where spatial units on the map have independently assigned “B” with probability p , or “W” with probability $1 - p$. We use free sampling when p is a priori given or known (Cliff and Ord, 1973). The second approach assumes non-free sampling, where each spatial unit has the same probability to be black or white, with general constraint that there are n_B black units, and n_W white units, and $n_B + n_W = n$. We rely on this approach when p is unknown and has to be estimated by $p = n_B/n$. Moments of JCS’s used at these approaches are not trivial for calculation and were derived by Cliff and Ord (1981), especially when $k > 2$. The last approach for testing the null hypothesis (lack of spatial autocorrelation) is permutation: we permute s times colours on the map and calculate the pseudo p -value.

JCS shows interesting similarities to case-control statistics developed in point pattern analysis (PPA), although the spatial structure is conceptualised in a different way (Waller, 2010). In PPA the control locations are a sample of all the possible control locations, whereas JCS considers all the possible control locations in the study-area (Anselin and Li, 2019, p. 192). The former is conceptually based on the assumption of spatial randomness of the events, the latter on a *lattice* perspective. For this reason, JCS is particularly suitable for areal entities, which leave no part of the study area unassigned (exhaustive tessellation) (Boots, 2008). The conceptual distinction between point patterns and areal entities includes neighbour relationships as well. In fact, spatial interaction in a point pattern is based on the closeness of the points, whereas areal data can also have spatially contiguous (adjacent) neighbours.

JCS is widely used in different research fields, from plant ecology to image analysis, but as pointed out above, its use in archaeology is extremely rare. Two works are worth mentioning: the pioneering paper published by Voorrips and O’Shea (1987), where the method is introduced in archaeological science for the first time; and the much more recent book chapter authored by Crema (2015), where a spatio-temporal version of JCS is applied. However, being a global test, its use in archaeology is often overlooked and relegated to a preliminary exploratory role. As a matter of fact, it is the identification of local outliers or clusters that enable significant spatial processes to be inferred. Furthermore, even if global spatial autocorrelation is not detected for a spatial dataset, local spatial autocorrelation might still occur (Ord and Getis, 2001). Therefore, a local version of JCS is necessary to foster its use in intra-site and landscape-scale archaeological research.

Table 1

All the possible combinations of local composition and configuration in Boots' LICD.

Types	Composition	Configuration
Hot Clump	$Pr(X \geq x) < 0.05$	BB significant
Hot Only	$Pr(X \geq x) < 0.05$	–
Clump Only (H)	–	BB significant
Cold Clump	$Pr(X \leq x) < 0.05$	WW significant
Cold Only	$Pr(X \leq x) < 0.05$	–
Clump Only (C)	–	WW significant
Dispersed	–	BW significant
No clusters	–	–

3. Local indicators for categorical data

The first mathematical formalisation of local measures of spatial association for categorical data was provided by Barry Boots. In the original study, Boots (2003) suggested the use a moving window of user-set size, usually $m \times m$ with r elements, centred on each cell, for analysing both *composition* and *configuration* of local spatial dependency.

Local composition describes whether the proportion of B cells in the moving window is significantly higher than in the entire region. If X is the random variable representing the number of B cells in the window and p_B is the proportion of B cells in the whole region, then the probability of having x cells is:

$$Pr(X = x) = \binom{r}{x} p_B^x (1 - p_B)^{r-x}, \quad (1)$$

where $x = 0, 1, \dots, r$.

When $Pr(X \leq x) < 0.05$ there is significant absence of B in r cells; when $Pr(X \geq x) < 0.05$ there is significant presence of B in r cells. These probabilities, in turn, can be derived from the previous binomial distribution, Eq. (1), as shown in the following equations:

$$Pr(X \geq x) = \sum_{h=x}^r \binom{r}{h} p_B^h (1 - p_B)^{r-h},$$

$$Pr(X \leq x) = \sum_{l=0}^x \binom{r}{l} p_B^l (1 - p_B)^{r-l}.$$

This method enables the assessment of local composition.

By estimating local configuration, the following question is addressed: for a given number of black cells x , does the number of joins between them in the r -cells window differ from the expected? Using rook's count (i.e. common edge neighbourhood), BB, WW, BW JCS are calculated for each window and statistical significance is derived in two ways: (1) using a normal approximation of join counts, when the number of cells in the window is $r \geq 30$, and the global ratio of B and W cells is $p_B \geq 0.2$ or $1 - p_B \geq 0.2$ (Cliff and Ord, 1981), (2) completely enumerating the distribution of join counts, or approximating it through a random sample of all possible outcomes (Boots, 2003).

Ultimately, local composition and local configuration are combined. Significant local composition cells are classified as “hot” (if there is a significant number of black cells in their window) or “cold” (if there is a significant number of white cells in their window). Local decomposition classifies the set of cells into clumps (when BB or WW is significant) or dispersed (when BW is significant). The combination of these two LICD measures produces eight possible classes (Boots, 2003) (see Table 1).

This procedure is used if the data do not show any global spatial autocorrelation. If autocorrelation exists, local composition can be affected, and local statistics are “too liberal” in identifying local spatial dependence. The procedure for such case is precisely described in Boots (2003).

The use of a regular ($m \times m$) moving window and the use of “clumps” (groups of B cells sharing an edge) as a key indicator of local composition, suggest that this method is primarily applicable to regular lattice, as acknowledged by the author himself (Boots, 2003,

p. 156). Although Anselin and Li (2019, p. 190) stated that Boots' method does not readily generalise to irregular layouts, there is no conceptual obstacle to extending this approach to irregular lattices. The only implication is that, necessarily, the windows will have different sizes. Also, if the windows are defined in terms of lags, first order window will be smaller in size than the smallest regular window used for regular lattice (3×3), and it would be less likely to find unusual configurational characteristics (Boots, 2006, p. 21). Therefore, for irregular lattice, higher order windows are required, and cell classification becomes more complex.

3.1. LICD for irregular lattices

Pietrzak et al. (2014) developed a simple method to estimate local spatial dependency for cells with variable size (Polish NUTS regions), by applying JCS to contiguity matrices (or k -nearest neighbour matrix) determined for each region. The advantage of this method over Boots' method is that it does not rely on a regular moving window and it is less computationally demanding (for large areas). The downside is that the statistical properties of this local test are unknown and this prevents statistical validation of local spatial dependences. Bivand et al. (2017) recently provided a revised implementation of the method. Their new approach is in fact a modification of Boots' LICD and replaces the Boots' moving window with symmetrical matrices of neighbours. Bivand et al. (2017) LICD was used to analyse the local spatial association of population pyramids in CEE regions. The authors identified 4 classes of pyramids, presented on a four-coloured map. The method consists of five consecutive steps:

1. For each region i they constructed the set of neighbours $N(i) = j$.
2. They calculated the probability p_k that a given region belongs to class $k \in \{1, 3A, 3B, 4\}$, or global composition.
3. For each region i they calculated the probability that within the neighbours of i (and given i) there are x_i regions of the same class as region i using a transformation of Boots formula: $P(X = x_i) = \binom{n_i}{x_i} p_k^{x_i} (1 - p_k)^{n_i - x_i}$. If $P(X \geq x_i) < p$ the region i is part of a significant cluster. If $P(X \leq x_i) < p$ it is part of a significant outlier. Sidák correction was applied for p . This provides local composition.
4. They calculated local JCS, using observation-wise weight objects. Significant BBs are described as possible clusters, BWs as dispersed. This provides local configuration, following type-1 measure as listed by Boots (2006).
5. The last step was the introduction of typology of regions: (1) *cluster* (significant number of B's and significant BB); (2) *clump* (only significant number of B); (3) *significant outlier* (significant number of non-B); (4) *dispersed* (significant BW); (5) *outlier in a heterogeneous area* (significant outlier with significant BW).

Three of the five types of regions were identified in their study area (see Fig. 2).

4. Application of LICD in archaeology

In this section, Boots' LICD and its revision for irregular layouts are tested on areal archaeological data. Code and data used for the analysis, as well as the outputs produced, have been made freely available to enable full reproducibility (Carrer et al., 2020). Implementation of the methods and discussion of the results are preceded by an analysis of the data type and their spatial structure.

4.1. Areal data in archaeology

Many archaeological objects can be easily approximated to point features in a map, like archaeological sites or artefacts/ecofacts distributed within an archaeological excavation. This has contributed to the success of spatial point pattern analysis (PPA) in archaeology,

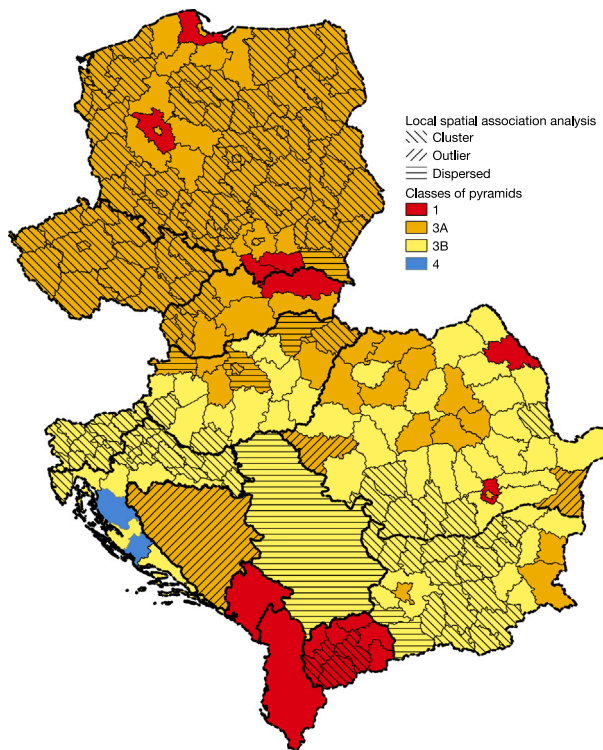


Fig. 2. Bivand et al. (2017) approach to local spatial association.

with a rapid acceleration in the last decade (Bevan et al., 2013). Less attention, instead, has been paid to areal data, although they are by no means less common in archaeological research. Two main types of areal entities are used in archaeology, at different scales: fieldwork grids and polygonisation of spatial features.

Grid systems are widely used to plan systematic field surveys, as well as to record the approximate location of artefacts, ecofacts and samples during the stratigraphic excavation of archaeological contexts.¹ The size of each grid cell varies from several square metres in a survey to 1 m² or under in the case of excavation. Conventionally, all the cells of the grid are squares (or rectangles) of the same size, to facilitate the manipulation and visualisation of spatial data. The application of spatial statistics on archaeological grids is well established and it includes the analysis of spatial association of numerical parameters (e.g. counts of artefacts) in neighbouring cells (Barceló and Maximiano, 2013). To the best of our knowledge, JCS or LICD methods have never been used to investigate archaeological grids.

Different GIS-based approaches in archaeology record spatial features using polygon geometries, but most of them register individual features (historical monuments, archaeological sites, protected areas), isolated in an unassigned background. Only Historic Landscape Characterisation (HLC) and similar characterisation approaches produce an exhaustive tessellation of each study area (Herring, 2009). HLC is a methodology developed in the UK by English Heritage (now Historic England) in the 1990s, to record how landscape character has changed over time and assess the persistence of historical aspects on current landscapes (Turner, 2006). HLC starts with the identification of a series of landscape character types that provide a simplified representation of the local landscape; these types are then identified in the study area and mapped as GIS polygons; the character type corresponding to each

polygon and the temporal variation of this type within each polygon are schematically recorded in the associated database. The outcome of HLC is a polygon layer (with attributes) that covers the whole study area, without gaps (Dabaut and Carrer, 2021). GIS queries enable the visual assessment of HLC datasets, but no attempts have been made to provide a more formal quantitative evaluation. The test presented in this paper is the first application of spatial statistics to HLC.

4.2. Spatial weights

The definition of neighbours and the attribution of spatial weights are important prerequisite for reliable statistics. These, in turn, depend on the spatial structure of areal data, as well as on the non-spatial information recorded.

Orientation, edges and cell size of archaeological grids, for example, are often arbitrary and do not necessarily reflect the spatial configuration of the archaeological context (its shape, its internal organisation, etc.). Spatial autocorrelation, in this case, might indicate the occurrence of continuous underlying processes (recorded as discrete events in the grid) or might be a consequence of the spatial inhomogeneity of the investigated context (hidden behind the regular tessellation). These underlying spatial processes and the regular lattice structure of the spatial dataset suggest that spatial proximity is the key parameter for the identification of neighbours in archaeological grids.

On the other hand, HLC data offer different challenges. The boundaries between HLC spatial entities are not arbitrary, they are assumed to represent actual discontinuities in the spatial variation of landscape character. This implies that the information recorded is inherently discrete and that local autocorrelation might reveal undisclosed spatial dependencies between landscape character types. However, user errors, inadequate cartographic resources, inconsistent approaches to polygonisation and uncertainty in the identification of clear spatial discontinuities might affect the reliability of these boundaries (Dabaut and Carrer, 2021). Therefore, the assessment of spatial association cannot be limited to contiguous neighbours, but needs to be extended to non-adjacent spatial entities. This can be achieved by considering higher-order neighbours (neighbours of neighbours) alongside first-order contiguous neighbours. The number of orders can be calibrated according to observed or inferred neighbour relations.

The identification of neighbours is followed by the definition of weights for the selected set of neighbours. Weights for each areal entity are usually standardised, so that their sum is 1. However, this method implies a reasonable understanding of the underlying spatial process. When such understanding cannot be realistically assumed, binary weighting (weight for each areal entity is equal to the number of its neighbours) is preferable.

4.3. Irregular lattice: Historic landscape characterisation

The HLC of Devon has been selected for testing LICD at landscape scale. Developed by English Heritage in the early 2000s (Turner, 2007), it is one of the first HLC produced in GIS and the raw dataset is freely available from the Archaeological Data Service (archaeologydataservice.ac.uk). Devon is a largely rural county in South-West England and its landscape is the product of a long history of human–environment interaction that goes back to the Neolithic (Ryder, 2013). The HLC of Devon records numerous character types of historic origin (primarily mediaeval and post-mediaeval) still surviving and visible on today's landscape (Fig. 3). The spatial distribution of mediaeval, post-mediaeval and modern character types in Devon has been identified as a suitable and interesting issue to be addressed using LICD. Local autocorrelation assessment can shed new light on the heterogeneous processes that determined the formation of this notorious rural landscape.

Devon's HLC is a very large dataset, consisting on more than 49,000 polygons (covering 6700 km²) associated with a complex array of non-spatial information. Besides, the geographical setting of the county,

¹ We have excluded from our overview and tests the grid datasets produced through geophysical prospections, since they are generally managed and manipulated using specialist tools and tailored methods.

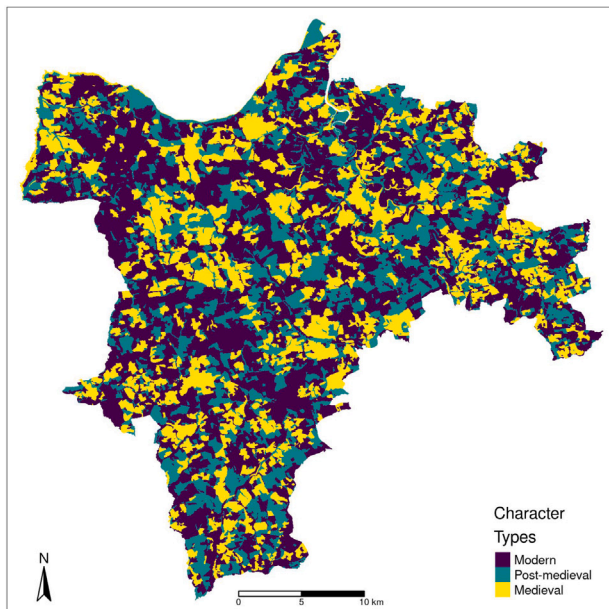


Fig. 3. HLC map representing the chronology of each character type in the current landscape of Torridge.

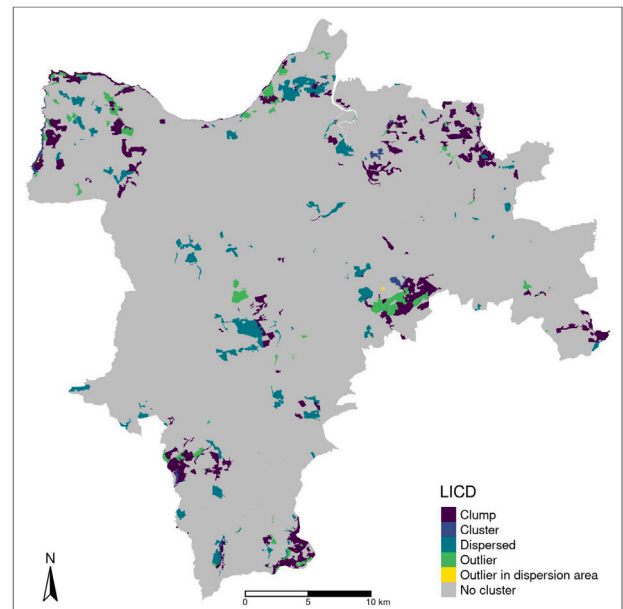


Fig. 4. Map of the LICD results for Torridge HLC.

and the coexistence of different land management systems, accentuate the heterogeneity of the investigated landscape. To reduce the computational power required for the analysis and manage the complexity of the landscape structure, a subsample of the study area has been selected for this test: the Torridge district. HLC shows that a significant portion of the current landscape in this district maintains a historic character, which makes it an ideal case-study for unravelling the correlation between the spatial distribution of character types and their chronology.

Global and local spatial association has been estimated for the three-colour HLC map of mediaeval, post-mediaeval and modern character types in the Torridge district. Regions (i.e. polygons) of unknown origin have been classified as mediaeval, as they represent vestiges of earlier landscape character. Modern regions correspond to the 51.7% of the entire dataset, post-mediaeval regions to the 35.2% and mediaeval regions represent the 13.1%. The reasonable balance between these chronological classes (none is rare or strongly predominant) enhances the likelihood of statistically meaningful results (Boots, 2006, p. 3). Three orders of contiguity neighbours have been considered to calculate binary spatial weights. This is the lowest order that enables the assessment of neighbour relations among non-proximal HLC regions. It is likely to be a conservative estimation, but a detailed investigation of the nature of each character type would be necessary to suggest more appropriate lags.

JCS has been performed for each class and each order of neighbours, using `joincount.multi` in the R package `spdep` (Bivand and Wong, 2018). The results display complex spatial associations of same-class and different-class neighbours (Table 2). None of the JCS is significant ($p < 0.01$) for three orders of neighbours, two are significant for two orders, two for one order, and two are not significant for any order. The inconsistency of neighbour relations is likely to depend on the non-stationarity of local processes, which can be addressed through LICD statistic. Therefore, the underlying global (landscape-scale) process is assumed to be stationary.

LICD has been applied to Torridge HLC by adapting the protocol developed in Bivand et al. (2017). The output of this statistic is a map of the typology of regions (polygons) as defined in the description of LICD methods. As expected, complex local association patterns have

been recorded, although most of the regions show no significant autocorrelation (Fig. 4). In order to facilitate the interpretation, maps of local association for each class have been produced and are provided in *Supplementary Figure 1*. An interactive visualisation of the LICD results and the original Torridge HLC dataset, created using the `mapview` package (Appelhans et al., 2020) is also available as *Supplementary Map*.

Mediaeval and post-mediaeval regions produce significant “clumps” and “clusters” in different sectors of Torridge. The former along the coast and near the NE border of the district. The latter on the SE and the SW border of the district. Post-mediaeval “dispersed” character types and “outliers” are well represented across the whole area, while they are much more infrequent among mediaeval regions. A large number of modern regions are “outliers” and “dispersed”, which indicates that they are surrounded by less modern regions than expected (under CSR) and have more non-modern regions among their neighbour than expected. Nevertheless, some small “clusters” and “clumps” of modern regions occur in the S/SW, NW and in the SE sectors of the district. These preliminary inferences clearly show that various processes of landscape change coexisted in different periods at different scales and a thorough analysis of LICD results can shed new light on the driving forces behind them.

4.4. Regular lattice: Barmose I excavation grid

Boots’ original LICD approach (Boots, 2003, 2006) has been used to test local autocorrelation in an archaeological grid. The dataset selected for this test comes from the Palaeolithic site of Barmose I, in Denmark. It was used by Blankholm (1991) to illustrate the application of different spatial analysis methods in archaeology, and was recently included by Carlson into an R package of archaeological data (Carlson and Roth, 2018) created for his handbook on quantitative analysis in archaeology (Carlson, 2017). The record of Barmose I includes the count of *debitage* for each cell of the excavation grid, and location coordinates for artefact of different types. The *debitage* count is a numerical variable, hence more suitable for routine autocorrelation analysis (Moran’s *I* and LISA), whereas artefact locations have been used by Orton (2004) to showcase the application of PPA in archaeology. For this LICD test, the two datasets have been combined and the occurrence of one artefact type in each cell of the grid has been considered. Since cores are among the most represented artefacts (17%

Table 2

Join-Count statistics of Torridge HLC for three classes (Mod: Modern, PM: Post-mediaeval, Med: Mediaeval) and three orders of contiguity neighbours (decimals truncated).

Contiguity order	JCS	Value of JCS	Expectation	Variance	z-value	p-value
First	Mod:Mod	771.1242	839.0020	102.1755	-6.7151	0.9999
	PM:PM	345.6026	391.8997	59.9811	-5.9778	0.9999
	Med:Med	54.8068	53.1667	9.9739	0.5193	0.3017
	PM:Mod	1091.0597	1147.2649	139.3980	-4.7604	0.9999
	Med:Mod	514.6464	422.7309	65.6800	11.3415	< 0.0001
	Med:PM	365.7601	288.9355	48.6345	11.0161	< 0.0001
Second	Mod:Mod	839.4692	839.0020	30.9970	0.0839	0.4665
	PM:PM	418.0195	391.8997	18.0340	6.1506	< 0.0001
	Med:Med	71.3133	53.1667	2.9557	10.5550	< 0.0001
	PM:Mod	1091.4778	1147.2649	41.2917	-8.6816	1
	Med:Mod	425.2992	422.7309	19.6594	0.5792	0.2812
	Med:PM	297.4208	288.9355	14.4803	2.2298	0.0128
Third	Mod:Mod	824.5191	839.0020	18.0468	-3.4092	0.9996
	PM:PM	394.3071	391.8997	10.1479	0.7557	0.2249
	Med:Med	60.3602	53.1667	1.5691	5.7425	< 0.0001
	PM:Mod	1137.7512	1147.2649	21.8816	-2.0337	0.9790
	Med:Mod	435.0227	422.7309	10.8705	3.7281	< 0.0001
	Med:PM	291.0393	288.9355	7.8376	0.7514	0.2261

of the assemblage) and are considered a reliable marker of activity areas, presence and absence of cores have been recorded for each cell of the grid (Fig. 5). Cells have been classed as “1” (B) when they had one or more cores, “0” (W) when they had none (Supplementary Figure 2). Patterns of cells with and without cores can provide insights on the spatial organisation of the site and also contribute to assessing possible data collection biases associated with the excavation grid.

The ratio of B and W cells (respectively 38 and 69) is balanced enough to guarantee reliable results. Since the orientation and size of the grid do not necessarily match the spatial variability of the dataset (as suggested by the *debitage* distribution), first order contiguity neighbours seemed inadequate for the test. On the other hand, high-order neighbour relations cannot be reliably assumed without further investigations. Such limitations have led to a cautionary use of two orders of contiguity neighbours in this trial. Global JCS (Table 3) show significant spatial autocorrelation ($p < 0.01$) for first-order “core” and “no core” joins and no autocorrelation for any second-order joins. LICD has been used to assess whether this difference in neighbour relations might be influenced by local patterns of cells.

The results of LICD (categorised according to the classes listed in Table 1) indicate that most of the cells are not clustered and only a small number show significant patterns (Supplementary Figure 3). A synthesis map, combining LICD results with the distribution of cores/non-cores cells, is provided in Fig. 6. Cells with no evidence of significant spatial association are ignored. Six “core” and three “no-core” cells at the centre of the grid show a significant positive composition (“Hot”), suggesting that the number of “core” cells among their neighbours is statistically significant. In the same sector of the grid, three “core” cells show significant positive configuration (“Clump”), as they are surrounded by more “core” cells than expected. Only one “core” cell shows significant composition and configuration. The unusual concentration of “core” cells at the centre of the grid is easily justified by the presence of a central hearth, around which most of the daily activities were probably performed (Blankholm, 1991). The “no-core” cells, isolated within a “core”-dominated area, were probably located too close to the firing area. On the other hand, a group of “no-core” cells near the top-right corner of the grid (NE sector of the site) deserves further attention. Three cells appear to have an unexpectedly low number of “core” cells among their neighbours (“Cold”), five are surrounded by more “no-core” cells than expected (“Clump”). These local associations suggest that, under CSR, more “core” cells would be expected on this sector of the excavation. Another isolated “no-core Cold” cell is close to the bottom-right corner of the grid.

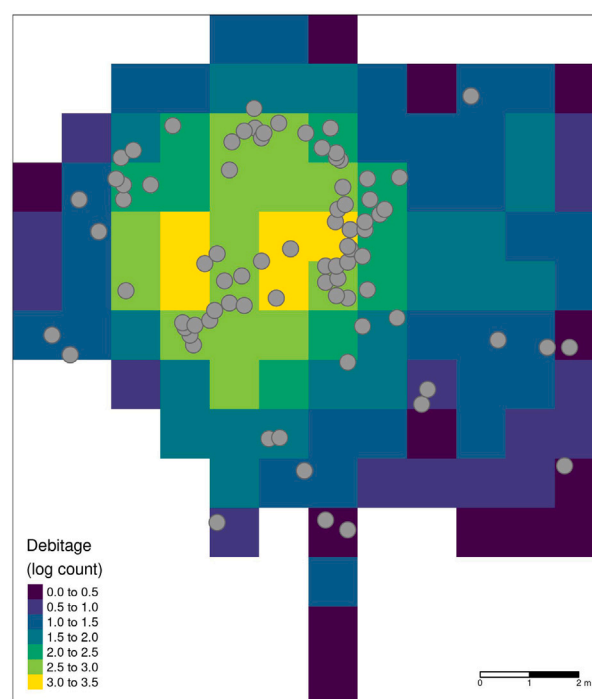


Fig. 5. Dispersion of debitage in the Barmose I grid (log scale) and location of flint cores within the excavation area.

5. Discussion

The application of LICD to Torridge HLC and the archaeological grid of Barmose I have revealed interesting and previously unrecognised spatial patterns.

The study of HLC data normally relies on expert knowledge and visual assessments and the identification of complex landscape-scale spatial relations is generally difficult through these informal approaches. In the case of Torridge, the inspection of HLC map (Fig. 3) does not provide any relevant information. Beyond an apparent clustering of each class, residual spatial complexity can be tentatively surmised, but it is almost impossible to disentangle without the support of spatial statistics. JCS results (Table 2) suggest complex spatial associations among and between the three classes, and LICD (Fig. 4) displays a considerable variability at local level. A detailed analysis of the landscape history of

Table 3

Join-Count statistics of cores presence/absence on Barmose I grid (C: Cores, NC: No-Cores) for two orders of contiguity neighbours (decimals truncated).

Contiguity order	JCS	Value of JCS	Expectation	Variance	z-value	p-value
First	C:C	8.9154	6.6320	0.4851	3.2783	0.0005
	NC:NC	24.0440	22.1320	0.5678	2.5372	0.0055
	C:NC	20.5404	24.7358	1.8256	-3.1049	0.9990
Second	C:C	7.7208	6.6320	0.3352	1.8805	0.0300
	NC:NC	22.0019	22.1320	0.4986	-0.1841	0.5730
	C:NC	23.7771	24.7358	1.1140	-0.9083	0.8181

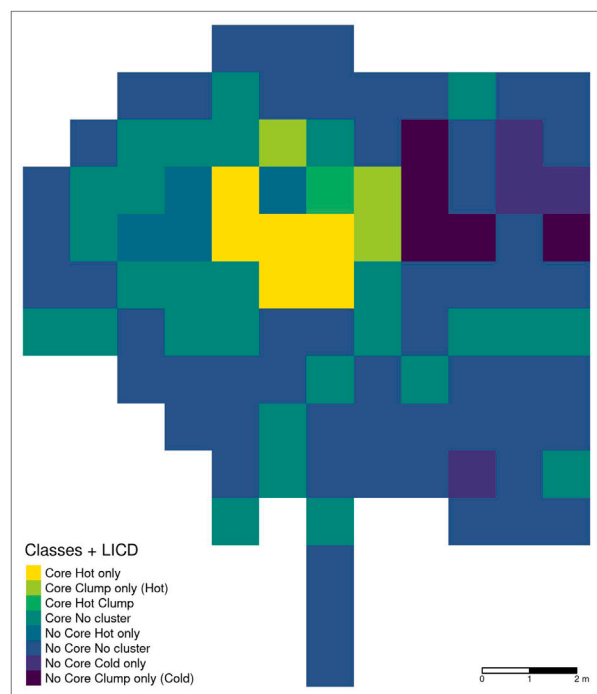


Fig. 6. Barmose I grid representing LICD results for each class (cores presence and cores absence).

Torridge would be necessary to explain the processes underlying such variability, but it would be beyond the scope of this paper. However, a rapid and general discussion is still required to assess the reliability of the results and evaluate the potential of LICD for landscape research. LICD maps show that modern and post-mediaeval regions are largely isolated among character types of different chronology. This suggests that mediaeval, post-mediaeval and modern types are more spatially interweaved than expected and that local landscape character is not strongly affected by the change occurring in neighbouring regions. In fact, it can even be argued that the transformation of some regions contributed to the preservation of the historic character in others. Such unpredicted spatial relations between landscape change and historic character can have important implications for future landscape management and development in Devon. If this seems to be the general trend in the study area, LICD recognises additional localised processes of landscape evolution. For example, mediaeval (or earlier) character types are clearly predominant in the coastal area. The scarce industrial and agricultural value of the coast might have contributed to the preservation of its pre-modern character, which is now protected as part of the local cultural and natural heritage. On the other hand, some parts of the district provide evidence of intense recent alteration. A large clump of modern regions in the South can be tentatively associated with the proximity to the A30 road, one of the major axes connecting London to the South West of England. Two clumps/clusters in the area around the villages of Merton and Clawton, indicating significant localised

change during the post-mediaeval period, deserve further historical and archaeological contextualisation.

Statistical analyses are routinely used to investigate archaeological grids, and LICD provides an additional tool for complementing and validating the results of other well-established quantitative methods. The analysis of core distribution within the excavation grid of Barmose I represents a solid assessment of the potential of LICD for intra-site spatial analysis. Local methods of point pattern analysis (PPA) showed that lithic cores are (like the other types) clustered around the hearth at the centre of the site (Orton, 2004), and object density decreases towards peripheral areas of the site. This pattern matches quite well the spatial distribution of *debitage* counts in each cell of the excavation grid (Fig. 5). LICD contributes to the spatial analysis of Barmose I through the identification of significant gaps in the distribution of cores. By investigating presence and absence of objects in each cell of the grid, LICD can detect empty cells that are closer to other empty cells than expected, or that are significantly dispersed among non-empty cells. LICD reveals that the spatial association of “core” and “no core” cells is generally compatible with CSR, except for two sectors of the site: (1) a high concentration of “core” cells close to the hearth location; (2) a high concentration of “no-core” cells in the NE sector of the site (Fig. 6). If the former simply backs the results of PPA, the latter provides new insights into the spatial organisation of cores. The unexpected lack of cores in the NE periphery of the excavation grid might mirror different first-order effects: the non-stationary distribution of activity areas within the site, localised taphonomic processes, a local bias in cell-based data recording, or a combination of all these processes. LICD has disclosed a previously unknown spatial pattern, opening new perspectives for the analysis of Barmose I object distribution.

6. Conclusions

This paper presents the first archaeological applications of Local Indicators for Categorical Data, a new tool to assess spatial autocorrelation of non-numeric variables for each spatial unit in irregular layouts (spatial datasets with regions of different sizes) and regular lattices (grid data).² In order to address the different spatial structures of these areal data, two different approaches were developed. The original Boots (2003) LICD was used for grid data, and a revised version proposed by Bivand et al. (2017) was adapted to irregular layouts. These methods were implemented in R, and tested on the HLC of the Torridge district (Devon, UK) and the excavation grid of Barmose I (Denmark).

Boots' LICD can be employed to investigate presence and absence of a selected variable within each cell of an archaeological grid, to identify local aggregation or dispersion patterns that might suggest the existence of specific spatial processes. It is worth pointing out that the information produced by LICD complements, rather than duplicating, the inferences produced by point pattern analysis. The latter detects spatial interaction at event locations, the former (based on a lattice perspective) identifies spatial patterns of events *and* non-event locations. The presence of unexpected “gaps” in the dataset, indicated

² Although there are no conceptual limitations to its use for point patterns, other methods are already available to address local association of multitype point data (Anselin and Li, 2019; Leslie and Kronenfeld, 2011).

by non-event cells that are closer or more dispersed than expected, might mirror the existence of hidden spatial processes that are worth investigating in detail. This bivariate assessment can also be undertaken for different types of co-occurring events, thus expanding the use of LICD beyond the presence/absence case. Possible applications of Boots' LICD in archaeology include the distribution of object types (similar to the example explored in this paper) or categorisation of artefact/ecofact tallies within excavation or survey grids, as well as the spatial variation of categorical parameters measured at evenly-spaced sampling locations (e.g. geoarchaeological sampling).

The revised LICD approach, based on symmetrical matrices of neighbours, can be used to test the autocorrelation of archaeological information recorded in spatial units of variable size. HLC and similar characterisation approaches (like HLA) have been identified as the main archaeological recipients of this method. Not only landscape characterisations produce irregular regions, but the information recorded for each region is primarily categorical: character type, historical origin of the type (addressed in this paper), dominant boundary class, etc. As mentioned above, spatial patterns of these categories are generally eyeballed using thematic maps and no formal analysis is implemented to provide scientific validation. The example presented in this paper suggests that LICD might help to exploit the full informative potential of HLC, by verifying visible patterns and disclosing hidden spatial relations. Furthermore, the identification of localised spatial processes will enable a finer-grain analysis of the driving forces that influence landscape change (Bürge et al., 2017). More reliable and higher-resolution inferences will boost the use of landscape characterisation for informing landscape development and management policies (Fairclough et al., 2018). On the other hand, the datasets produced with HLC and comparable methods are not the only types of archaeological information suitable for LICD applications. Other examples include survey data recorded within each field (or cadastral parcel) rather than within standardised survey units, or maps of historic settlements reporting chronological, functional or ownership information for each dwelling unit.

In conclusion, the results of the two tests described above suggest that LICD will significantly contribute to the advancement of spatial analysis in landscape and intra-site archaeology. It is hoped that this paper will promote its use in archaeological research and heritage management.

CRediT authorship contribution statement

Francesco Carrer: Conceptualisation, Methodology, Software, Writing - review & editing. **Tomasz M. Kossowski:** Software, Writing - review & editing. **Justyna Wilk:** Software, Review. **Michał B. Pietrzak:** Software, Review and editing. **Roger S. Bivand:** Software, Validation, Review and editing.

Declaration of competing interest

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

Acknowledgements

We are very grateful to the two anonymous reviewers for their useful comments. F. Carrer's research is supported by the Newcastle University Academic Track Fellowship programme (NUAct). Figs. 3–6 are produced using the R library **tmap** (Tennekes, 2018).

Appendix A. Data and code

Original datasets and R code for this paper can be found at <https://doi.org/10.5281/zenodo.4283767>. A CODECHECK certificate for this paper is available at <https://doi.org/10.5281/zenodo.4279275>. CODECHECK is an open-science initiative to facilitate sharing of computer programmes and results presented in scientific publications (codecheck.org.uk).

Appendix B. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jas.2020.105306>.

References

- Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. Kluwer, Dordrecht.
- Anselin, L., 1995. Local indicators of spatial association-LISA. *Geogr. Anal.* 27 (2), 93–115.
- Anselin, L., Li, X., 2019. Operational local join count statistics for cluster detection. *J. Geogr. Syst.* 21 (2), 189–210.
- Appelhans, T., Detsch, F., Reudenbach, C., Woellauer, S., 2020. Mapview: Interactive Viewing of Spatial Data in R. R Package Version 2.9.4. <https://cran.r-project.org/package=mapview>.
- Bai, H., Li, D., Ge, Y., Wang, J., 2016. Detecting nominal variables' spatial associations using conditional probabilities of neighboring surface objects' categories. *Inform. Sci.* 329, 701–718.
- Barceló, J.A., Maximiano, A., 2013. The mathematics of domestic spaces. In: Madella, M., Kovacs, G., Berzsenyi, B., Briz i Godino, I. (Eds.), *The Archaeology of Household*. Oxbow Books, Oxford, pp. 6–22.
- Bevan, A., Crema, E., Li, X., Palmisano, A., 2013. Intensities, interactions and uncertainties: Some new approaches to archaeological distributions. In: Bevan, A., Lake, M. (Eds.), *Computational Approaches To Archaeological Space*. Left Coast Press, Walnut Creek, pp. 27–52.
- Biagetti, S., Alcaina-Mateos, J., Crema, E.R., 2016. A matter of ephemerality: the study of Kel Tadrart Tuareg (southwest Libya) campsites via quantitative spatial analysis. *Ecol. Soc.* 21 (1).
- Bivand, R.S., Wilk, J., Kossowski, T., 2017. Spatial association of population pyramids across Europe: The application of symbolic data, cluster analysis and join-count tests. *Spat. Stat.* 21 (B), 339–361.
- Bivand, R.S., Wong, D.W.S., 2018. Comparing implementations of global and local indicators of spatial association. *Test* 27 (3), 716–748.
- Blankholm, H.P., 1991. *Intrasite Spatial Analysis in Theory and Practice*. Aarhus University Press, Aarhus.
- Boots, B., 2003. Developing local measures of spatial association for categorical data. *J. Geogr. Syst.* 5 (3), 139–160.
- Boots, B., 2006. Local configuration measures for categorical spatial data: binary regular lattices. *J. Geogr. Syst.* 8 (1), 1–24.
- Boots, B., 2008. Tessellation. In: Kemp, K.K. (Ed.), *Encyclopedia of Geographic Information Science*. SAGE Publications, Thousand Oaks, CA.
- Brandolini, F., Carrer, F., 2020. Terra, silva et paludes. Assessing the role of alluvial geomorphology for Late-Holocene settlement strategies (po plain – N Italy) through point pattern analysis. *Environ. Archaeol.* 1–15. <http://dx.doi.org/10.1080/14614103.2020.1740866>.
- Bürge, M., Bieling, C., von Hackwitz, K., Kizos, T., Lieskovský, J., Martín García, M., McCarthy, S., Müller, M., Palang, H., Plieninger, T., Printsmann, A., 2017. Processes and driving forces in changing cultural landscapes across Europe. *Landsc. Ecol.* 32 (11), 2097–2112.
- Carlson, D.L., 2017. *Quantitative Methods in Archaeology Using R*. Cambridge University Press, Cambridge.
- Carlson, D.L., Roth, G., 2018. Archdata: Example datasets from archaeological research. R package version 1.2 <https://CRAN.R-project.org/package=archdata>.
- Carrer, F., 2017. Interpreting intra-site spatial patterns in seasonal contexts: an ethnoarchaeological case study from the Western Alps. *J. Archaeol. Method Theory* 24, 303–327.
- Carrer, F., Kossowski, T.M., Wilk, J., Pietrzak, M.B., Bivand, R., 2020. rsbivand/LICD.article: Release of paper reproduction code, v1.0.0. <http://dx.doi.org/10.5281/zenodo.4283767>.
- Cliff, A.D., Ord, J.K., 1973. *Spatial Autocorrelation*. Pion, London.
- Cliff, A.D., Ord, J.K., 1981. *Spatial Processes: Models and Applications*. Pion, London.
- Crema, E.R., 2015. Time and probabilistic reasoning in settlement analysis. In: Barceló, J.A., Bogdanovic, I. (Eds.), *Mathematics and Archaeology*. CRC, pp. 314–334.
- Dabaut, N., Carrer, F., 2021. Historic landscape characterisation: technical approaches beyond theory. *Landscape* (in press).
- Epperson, B.K., Alvarez-Buylla, E.R., 1997. Limited seed dispersal and genetic structure in life stages of *Cecropia obtusifolia*. *Evolution* 51 (1), 275–282.

- Fairclough, G., Herlin, I.S., Swanwick, C., 2018. Landscape character approaches in global, disciplinary and policy context: introduction. In: Fairclough, G., Herlin, I.S., Swanwick, C. (Eds.), *Routledge Handbook of Landscape Character Assessment*. Routledge, London, pp. 3–20.
- Franklin, J., Wejnert, K.E., Hathaway, S.A., Rochester, C.J., Fisher, R.N., 2009. Effect of species rarity on the accuracy of species distribution models for reptiles and amphibians in southern California. *Divers. Distrib.* 15 (1), 167–177.
- Geary, R.C., 1954. The contiguity ratio and statistical mapping. *Incorporated Stat.* 5, 115–145.
- Getis, A., 2008. A history of the concept of spatial autocorrelation: A geographer's perspective. *Geogr. Anal.* 40 (3), 297–309.
- Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. *Geogr. Anal.* 24, 189–206.
- Hansen, M.H., Nair, V.N., Friedman, D.J., 1997. Monitoring wafer map data from integrated circuit fabrication processes for spatially clustered defects. *Technometrics* 39 (3), 241–253.
- Herring, P.C., 2009. Framing perceptions of the historic landscape: Historic landscape characterisation (HLC) and historic land-use assessment (HLA). *Scott. Geogr. J.* 125 (1), 61–77.
- Kossowski, T., 2018. Spatial dependence: how can we identify and measure it? In: Kuttar, D. (Ed.), *Visegrad Mosaic – New Colours and Old Contours. Observing and Understanding the Spatial Features of Socio-Economic Processes in East Central Europe*. Miskolci Egyetem, Miskolc, Hungary, pp. 39–60.
- Kvamme, K.L., 1990. Spatial autocorrelation and the classic maya collapse revisited: Refined techniques and new conclusions. *J. Archaeol. Sci.* 17 (2), 197–207.
- Leslie, T.F., Kronenfeld, B.J., 2011. The colocation quotient: A new measure of spatial association between categorical subsets of points. *Geogr. Anal.* 43 (3), 306–326.
- López, F., Matilla-García, M., Mur, J., Ruiz Marin, M., 2010. A non-parametric spatial independence test using symbolic entropy. *Reg. Sci. Urban Econ.* 40, 106–115.
- Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27, 209–220.
- Meng, B., Wang, J., Liu, J., Wu, J., Zhong, E., 2005. Understanding the spatial diffusion process of severe acute respiratory syndrome in Beijing. *Publ. Health* 119 (12), 1080–1087.
- Moran, P.A.P., 1948. The interpretation of statistical maps. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 10, 243–251.
- Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. *Biometrika* 37 (17–23).
- Ord, J.K., Getis, A., 1995. Local spatial autocorrelation statistics: Distributional issues and an application. *Geogr. Anal.* 27, 286–306.
- Ord, J.K., Getis, A., 2001. Testing for local spatial autocorrelation in the presence of global autocorrelation. *J. Reg. Sci.* 41 (3), 411–432.
- Orton, C., 2004. Point pattern revisited. *Archeol. Calcolatori* 15, 299–315.
- Pietrzak, M.B., Wilk, J., Kossowski, T., Bivand, R.S., 2014. The application of local indicators for categorical data (LICD) in the spatial analysis of economic development. *Comp. Econ. Res.* 17 (4), 203–220.
- Premo, L.S., 2004. Local spatial autocorrelation statistics quantify multi-scale patterns distributional data: An example from the maya lowlands. *J. Archaeol. Sci.* 31 (7), 855–866.
- R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ryder, L., 2013. *The Historic Landscape of Devon: A Study in Change and Continuity*. Windgather Press.
- Tennekes, M., 2018. Tmap: Thematic maps in R. *J. Stat. Softw.* 84 (6), 1–39.
- Tobler, W., 1970. A computer movie simulating urban growth in the detroit region. *Econ. Geogr.* 46 (2), 234–240.
- Turner, S., 2006. Historic landscape characterisation: A landscape archaeology for research, management and planning. *Landscape Res.* 31 (4), 385–398.
- Turner, S., 2007. *Ancient Country: the historic character of rural Devon*. Devon Archaeological Society Occasional Paper, vol. 20, Devon Archaeological Society, Exeter, UK.
- Vaz, E., 2020. Archaeological sites in small towns - A sustainability assessment of northumberland county. *Sustainability* 12 (2018), 1–13.
- Voorrips, A., O'Shea, J.M., 1987. Conditional spatial patterning: Beyond the nearest neighbor. *Am. Antiq.* 52 (3), 500–521.
- Waller, L., 2010. Point process model and methods in spatial epidemiology. In: Gelfand, A.E., Diggle, P.J., Guttorp, P. (Eds.), *Handbook of Spatial Statistics*. CRC, pp. 403–423.