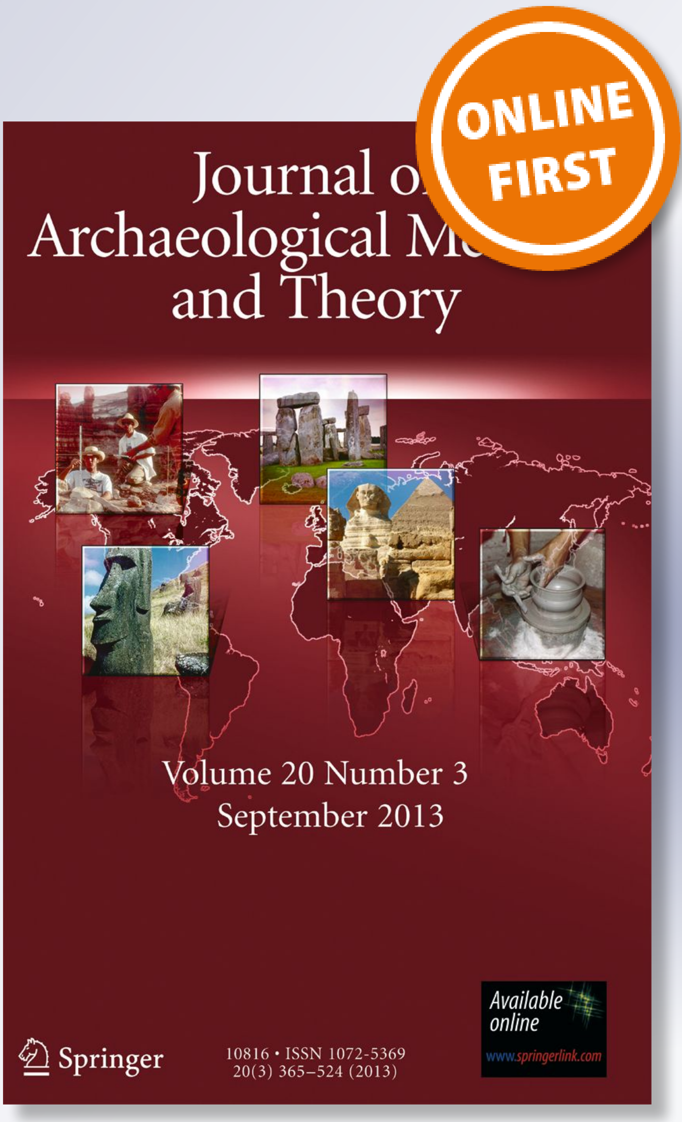
A Cost-Based Ripley's K Function to Assess Social Strategies in Settlement Patterning

Joan Negre, Facundo Muñoz & Juan Antonio Barceló

Journal of Archaeological Method and Theory

ISSN 1072-5369

J Archaeol Method Theory
DOI 10.1007/s10816-017-9358-7



J Archaeol Method Theory
https://doi.org/10.1007/s10816-017-9358-7

A Cost-Based Ripley's *K* Function to Assess Social Strategies in Settlement Patterning

Joan Negre[1] · **Facundo Muñoz**[2] · **Juan Antonio Barceló**[1]

Abstract Most quantitative approaches to distributional analysis in archaeology assume a homogeneous study surface that is amenable to easy generalisations. This framework has been widely used to describe settlement processes, disregarding the spatial heterogeneity inherent to geographic reality. In other words, researchers have often assumed that the correlation between the elements of a spatial distribution is a function of the Euclidean distance (*i.e.* straight line distance) between them. Other archaeological studies have tested alternative measures to Euclidean distances, such as cost-based ones, both to describe optimal routes and to assess spatial autocorrelation in a point pattern. Nevertheless, until now there has been no suitable model to introduce these measures into spatial statistical equations. In order to overcome this obstacle, we approach the implementation problem inversely by embedding the spatial pattern under study into a Euclidean frame of reference based on its cost-distance pairwise matrix. This paper describes the application of this methodology on one of the main tools used by archaeologists to assess settlement patterns: Ripley's *K* function. We present two case studies, covering both macroscale and mesoscale, with significant variations in the results depending on the use of the Euclidean or cost-based approach. Data, functions and results have been *R*-packaged for the sake of reproducibility and reusability, allowing other researchers to build upon our methods.

Keywords Ripley's *K* function · Cost distances · Settlement patterns · Spatial heterogeneity · Euclidean embedding · Multidimensional scaling

✉ Joan Negre
negreperez@gmail.com

[1] Laboratori d'Arqueologia Quantitativa, Departament de Prehistòria, Universitat Autònoma de Barcelona, Edifici B-Fac. Filosofia i Lletres, 08193 Barcelona, Bellaterra, Spain

[2] UR AGPF, INRA, 45610 Ardon, France

Published online: 09 November 2017

Introduction

The use of spatial statistics in archaeology can be traced back to the advent of geographic quantitative approaches over the past several decades (Matthews 1981; Fotheringham and O'Kelly 1989; Fotheringham *et al.* 2000; Rogerson 2002). These methods have provided a powerful toolbox for addressing many archaeological problems, but their implementation has mostly rested on a trial basis. In this context, different authors have developed approaches for analysing settlement patterns using different statistical methods. Most have focused on either the correlation between habitat and socioenvironmental features (such as physiographic characters, roads or irrigation canals, among others) or neighbourhood dependence between sites (Kvamme 1997; Bevan and Conolly 2006; Conolly and Lake 2006; Schwarz and Mount 2006; Mayer 2006; Verhagen et al. 2010; White and Surface-Evans 2012). Both approaches are interrelated in that they aim to analyse human occupation strategies and their spatial signatures. Nevertheless, these methodological advances have not always been integrated into a proper archaeological discussion (Barceló and Pallarés 1998; Church *et al.* 1999; Lock and Harris 2000; Woodman and Woodward 2002; Philipps 2004).

Our study focuses on one of the main tools used by archaeologists to assess neighbourhood dependence: Ripley's *K* function (Ripley 1976). This statistical test is widely used to detect clustering or dispersion trends, where the null hypothesis is that the point process under analysis is a homogeneous Poisson distribution. Its popularity is related to the limitations of previous tools, such as nearest neighbour and quadrant analyses—scale-dependent statistics that fail to detect second order processes. Ripley's *K* function aims to illustrate the underlying social rationale in the settlement pattern under study against the topography of the region of study, which is assumed to be an absolute plane (Fotheringham *et al.* 2000; Veen and Schoenberg 2006; Macchi 2009). Since we know that topographical reality is more complex than Euclidean space, the results of this statistic must be linked to a confusing mix of interactions between social and topographical sources of variability. The main limitation of the *K* function, therefore, is related to the assumption of an oversimplified Euclidean frame of reference, in which most models use a homogeneous and undifferentiated surface, amenable to easy generalisations (Bevan and Conolly 2006; Negre 2015a).

The spatial variation in archaeological elements should be analysed as primarily structured by complex social behaviour, as correlated to landscape variation and many other geological and ecological mechanisms that can vary widely across locations. That is, two settlements separated by a short Euclidean distance can be very far away in terms of social distance when there is a topographical barrier between them. Conversely, a long Euclidean distance does not necessarily preclude a social relationship if a road or other connection exists that reduces travel time from one site to the other. Likewise, even if the capital is relatively far away from a rural outpost, the flow of social interaction between them might be far greater that the social interaction with other villages that may be nearer in the Euclidean sense.

Therefore, what appears to be random in the Euclidean space might not be within a socially built territory. Consequently, it is not possible to test socially intentional spatial aggregation of settlements at particular places against a null hypothesis based on equally spaced points; it is necessary to consider the substantial heterogeneity of geographic reality.

In this paper, we propose a new way of applying Ripley's *K* function. Instead of presuming spatial homogeneity in the area of study, we assume a limited amount of aggregation and clustering that is not the consequence of intentional social activity, but the obvious result of the pressure of physical environment. We rely then on the calculation of an appropriate cost surface of the area under study, which accounts for the influence of the geographic criteria on the distribution of the habitat. That source should include both physical features (slope, rivers, geology, *etc*.) and the human interventions created to navigate them (mainly river crossings and roads) (Marble 1996; van Leusen 2002; Wheatley and Gillings 2002; Conolly and Lake 2006; Herzog 2014b). In these conditions, we will use Ripley's *K* function to characterise the degree of settlement clustering generated by intentional social activities, going well beyond the clustering that can be attributed to the particular topographical variation.

Our proposal enables more exact insight of the spatial structure of settlement patterns and its relationship with historical occupation dynamics, taking into account the geographic reality in which humans socialised and testing whether human decisions regarding settlement strategies were simply adapted to the territory or if they contributed to modify the territory in a particular way.

Methodological Proposal

The current algorithms for estimating the Ripley's *K* function take inputs of a set of coordinates that are assumed to be Euclidean. It is not possible to directly input a set of pairwise cost-based distances, which would require methodological developments beyond the scope of this paper. The problem is ultimately related to the development of a valid complete spatial randomness (CSR) null hypothesis. When the region under study is heterogeneous (Fig. 1a), we do not expect a uniform intensity for the underlying point process (*i.e.* CSR), which is based in the Euclidean assumption. This approach tries to codify that heterogeneity in a cost surface (Fig. 1b), allowing us to analyse the clustering pattern after accounting for this source of variability. With this in mind, we conceive the implementation inversely, modelling a new spatial distribution based on its cost-based pairwise matrix (Fig. 1c). Thanks to this approach, it is possible to offer an alternative clustering measure by embedding the point process under study into a Euclidean frame of reference (Fig. 1d and e) that is more representative of the geographic distances between its elements. The structure of this new spatial distribution is now only related to the social source of variability (*i.e.* human decisions about habitat location) for which Ripley's *K* function (Fig. 1f) can be estimated for further interpretation.

Alternative measures to Euclidean metrics have been widely tested in other disciplines. In archaeology, cost-based metrics are the most usual alternative for understanding the optimal paths between settlements and site catchment areas (Gaffney and Stančič 1991; Llobera 2000; Bell and Lock 2000; Ducke and Kroefges 2008; Herzog 2014b). They have also been used as a reference for spatial autocorrelation analysis when stationarity cannot be assumed (Negre 2015a; Negre *et al.* 2016b), that is, when the spatial point process cannot verify constant intensity and uniform correlation depending only on the lag vector between point pairs (Schabenberger and Gotway 2005; López-Quílez and Muñoz 2009; Møller and Toftaker 2012).

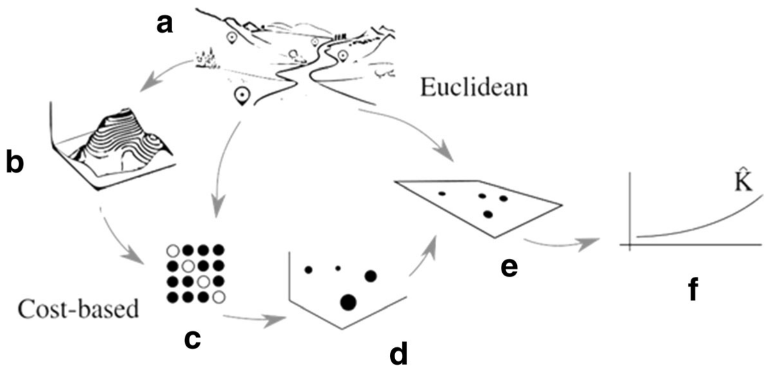


Fig. 1 Cost-based and Euclidean workflows. From a spatial pattern in a heterogeneous environment (a), a cost surface is derived (b). The matrix of cost-based distances (c) is computed and embedded into a Euclidean space (d) of possibly large dimensions. This point pattern is then projected into the most relevant two-dimensional Euclidean space (e), and the standard methods for estimating the Ripley's K function (f) are applied. The standard Euclidean workflow jumps directly to the bi-dimensional space (e), ignoring the geographic heterogeneity

From a methodological perspective, the main goal of cost distances is to define the least-cost path to reach a known point from each cell location in the original raster dataset. The measurement algorithms present the length of the irregular vectors formed by a spatial distribution using the shortest weighted distance, that is, the path with the least accumulated cost. The base where these calculations are performed is the cost surface; the purpose of which is to assign each cell of a raster layer with an impedance value that quantifies the ease with which it can be crossed (Soule and Goldman 1972). Researchers have a wide variety of options to calculate these friction surfaces, each dependent on their geophysical and social assumptions (Llobera and Sluckin 2007; Herzog 2014a, b; Negre 2015a). The resulting friction model can be defined as a function f, which describes for each cell of our territory a real, positive value representing the difficulty of crossing them, that is, its cost-weighted density. Therefore, the cost of a displacement dx at point x is $f(x)dx$, and the cost of any path in a region A can be calculated by integrating along the pathway (Muñoz 2012; López-Quílez and Muñoz 2009). More formally, the cost of a path $\alpha: (0, 1) \rightarrow A$ between points $s_1, s_2 \in A$ will be

$$\int_0^1 f(\alpha(t))\alpha'(t)dt$$

These measures verify the properties of non-negativity, symmetry and triangle inequality. Thus, we implicitly assume the hypothesis of movement isotropy, meaning that directional slope has a minor effect on the quality of the correlation model (Llobera 2000; van Leusen 2002).

Once the matrix of cost-based distances is created, a bi-dimensional point pattern that approximates the given pairwise distances can be derived by ordination techniques, such as multidimensional scaling (MDS). Specifically, these methods seek a Euclidean representation of a set of similarity relations between the objects. In this case, the similarities are represented in the form of an $n \times n$

symmetrical matrix of all pairwise cost distances, which represent the geographic distance between the settlements. Unlike other ordination methods, MDS makes few assumptions about the nature of the data, with no linear or modal relationships taken for granted in the data matrix. This makes it well suited for a wide variety of data; any distance measure between data points is allowed (Holland 2008). Moreover, because MDS operates on dissimilarities, no statistical distribution assumptions are necessary. The result constitutes a two-dimensional point pattern in the Euclidean space R^2, whose clustering can be measured using the standard form of the Ripley's *K* function that operates under this setting.

Package Overview

In what follows, we describe the tools used for each of the steps for the analysis. We packaged the data, functions and results for our case studies into an *R* package for the sake of reproducibility and reusability so that other researchers can build on our methods (Marwick 2016). The package is available at gitlab.com/famuvie/cbK (Muñoz and Negre 2017) and includes the convenience function *cbK*() for automatic calculation of all results given the cost surface and a point pattern (*i.e.* it automates the steps c to f in Fig. 1).

Computing the Cost Surface (Fig. 1b)

Slope is calculated by using Llobera and Sluckin's (2007) symmetric quadratic cost function. Soil properties are used as multipliers of cost expenditure for crossing different land covers (Soule and Goldman 1972). Swampy riverbanks, peat bogs and heavy brush areas are given a value of more than 1, and known road paths, below 1. Watercourses are likewise considered as multipliers, in the range of 5 to 10 for small streams and 15 for navigable rivers (Herzog 2014b). We have considered that displacement along the watercourse was not common, since the riverbanks are muddy, and fast-flowing rivers overflow periodically. Roads and strategic river crossings are placed where archaeological and written sources indicate. In the second case there are only small streams, with a value of 5 assigned to them. The resulting cost surface is box-plotted, and outliers (an order of magnitude higher) are filtered.

This step cannot be automated. It is up to the researcher to adequately represent the meaningful aspects of the problem as a cost-surface. The following steps are automatically taken care of by the function *cbK*() in our package and are described as follows for the sake of transparency and comprehensiveness.

Computing the Cost-Distance Matrix (Fig. 1c)

We leveraged the function *distmatGen*() from the package *geoRcb* (Muñoz 2015). For a given set of points and conductivity (*i.e.* inverse cost) surface, this function generates a cost-based distance matrix among the points. Behind the scenes, the function uses the package *gdistance* (van Etten 2015). The use of the conductivity rather than cost surface is more convenient computationally, especially when high or infinite costs are involved.

Multidimensional Scaling of Spatial Point Process (Fig. 1d and e)

Given the distance matrix from the previous step, we obtain an MDS representation that approximates the distances in R^2 using the R function: *cmdscale*(). We can quantify the precision of the approximation with the relative magnitude of the first two eigenvalues with respect to the total sum of eigenvalues (Cox and Cox 2000). This measure can be interpreted as the proportion of the geographic heterogeneities captured by the MDS approximation. It depends on both the topography and the point pattern.

The reduction of dimensionality generally entails some loss of information. However, it is the optimal two-dimensional Euclidean representation of the desired distance structure. This implies that it is always better or at least equivalent to using the standard methodology with the original Euclidean coordinates.

Computing Ripley's K Function and Monte Carlo Envelope for Significance Testing (Fig. 1f)

Finally, we use functions from the *spatstat* package to compute statistics K and Ld and to model Monte Carlo envelopes to test the CSR hypothesis (Baddeley and Turner 2015; Baddeley et al. 2015).

Comparing the Results with Euclidean Distances

For the sake of reference and comparison, we also compute K and L functions for the original point pattern using Euclidean distances and thus disregarding the geographic source accounted for in the cost-based analysis.

Case Study 1: Ancient and Medieval Rural Settlement Patterns in a Riverside Region

Study Problem and Materials

In this case study, we analyse and compare the rural settlement strategies from the Roman imperial period (2nd–3rd centuries) and the Caliphate of Cordoba (10th–11th centuries) within the territory of Tortosa (Tarragona, Catalonia). We chose those two phases for being representative of the thriving States (Roman Empire and Al-Andalus Caliphate) in the same region but separated by nearly a thousand years. The area of Tortosa (approximately 1700 km^2) coincides with the medieval boundaries established by the feudal administration in 1149 AD just after conquest. The lower course of the Ebro River was a central axis of the region, delimiting a geomorphologic unity identified as a tectonic depression with a northwest-east orientation. The width of the valley varies from 2 to 4 km, and the sedimentology indicates a silt composition from the fluvial terraces, formed throughout the Holocene. Most agrarian activity took place here, as the fertile soil enabled the production of abundant and good quality crops (Fig. 2).

Archaeological study uncovered about 5000 pottery fragments (800 MNI) as well as more than 1000 Latin and Arabic written texts, comprising both primary and edited

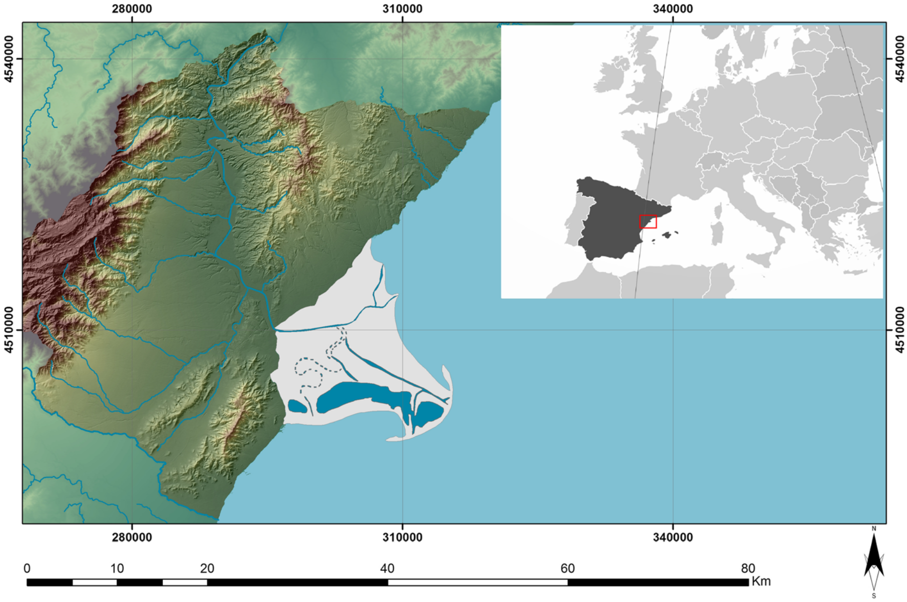


Fig. 2 Situation map

sources (Negre 2013, 2014, 2015b; Martí and Negre 2015). Twenty-nine settlements pertained to the ancient phase, while 37 correspond to the proposed medieval chronology. Analysing the occupation dynamics of this territory might shed light on the specific human strategies regarding settlement processes, becoming a precise indicator of social change once contrasted with written and other historical sources. State directives or guidelines and human strategies to exploit the rural environs tend to affect settlement patterns, and consequently, the transformations observed can be formally analysed as a result of sociopolitical decisions (Fig. 3).

Results

The first study period (Imperial Roman, 2nd–4th centuries) presents very different results depending on whether the approach is based on Euclidean or cost distances. While the first method points to a spatial trend within the margins of complete spatial randomness, the use of its cost-based counterpart allows us to detect a significant pattern of clustering at all scales. The same is true when analysing the medieval period (Islamic Caliphate, 10th–11th centuries). The Euclidean approach shows a very slight tendency to clustering starting at 5 km, while our approach indicates this trend more clearly beginning at 10,000 cost units (Fig. 4).

From a geographic viewpoint, the importance of the river stands out as a source of heterogeneity in spatial variation. Euclidean-based approaches wrongly assign an equal value to distances between sites on the same riverbank as to sites placed on opposite sides of the river. Without means to cross the river, it remains a formidable obstacle to communication.

Extensive research has addressed the issue of clustering in historic rural settlement patterns. Social mechanisms, such as those developed by states or other forms of social

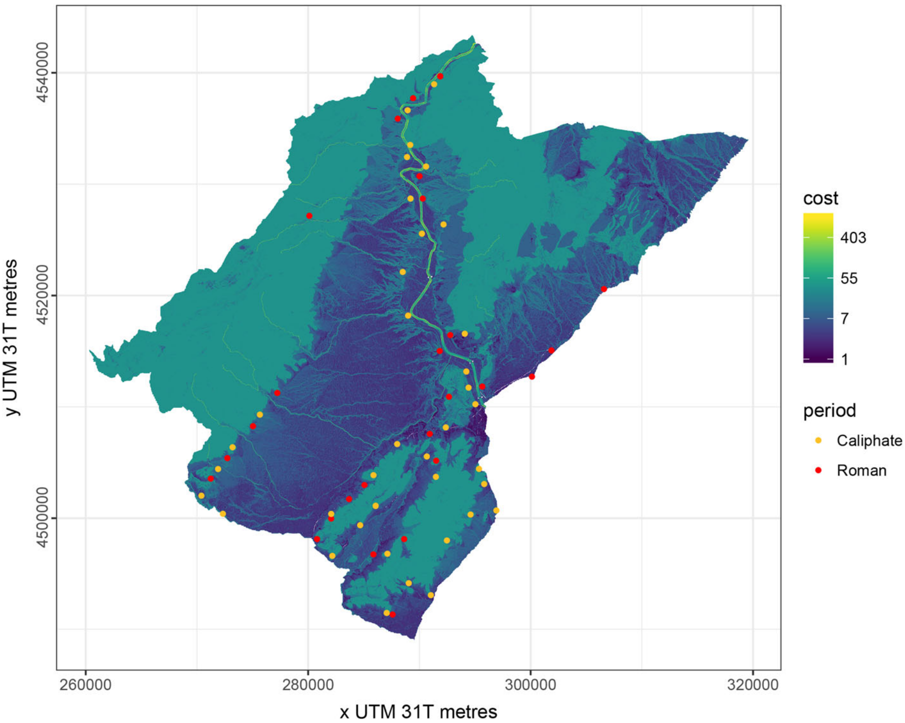


Fig. 3 Cost surface and sites in ancient and medieval phases

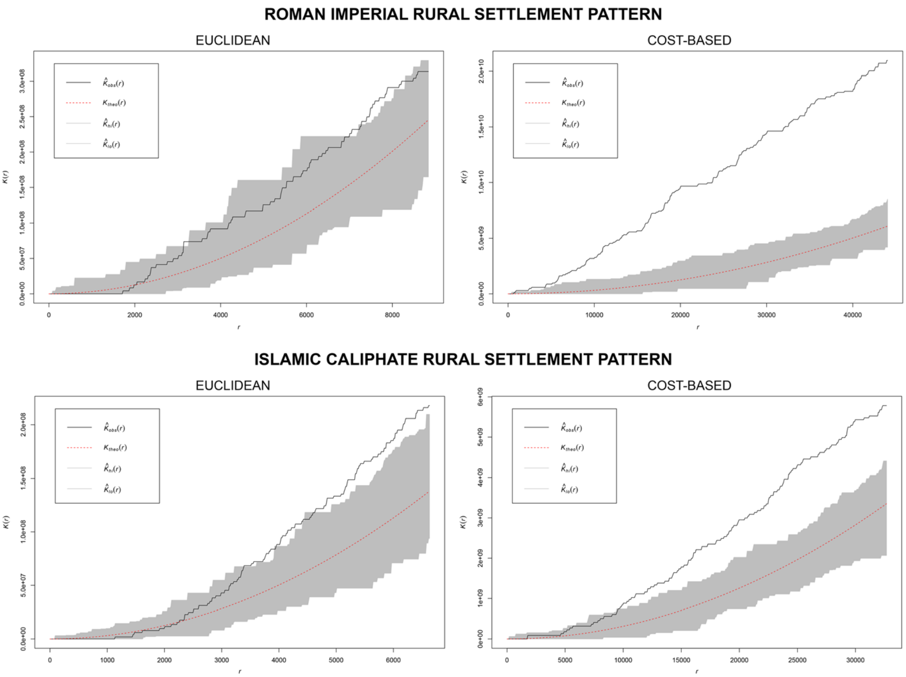


Fig. 4 Comparison between Euclidean and cost-based Ripley's *K* function

organisation, tend to act over settlement strategies within their borders in an extremely sound way. Different theoretical approaches to rural location processes have proposed three main stages in their development: colonisation, by which the occupied territory of a population expands; spread, through which settlement density increases with a tendency to short-distance dispersal; and competition, a process that produces a regularity in settlement pattern when there are sufficient rural dwellers to compete for space (Hudson 1969; Birch and Hudson 1970). The last stage is also considered an indicator of a fully mature settlement process on a certain scale, with settlements located at very similar distances from each other (Christaller 1966).

In summary, our results are concordant with the spreading phase described above, in which settlement multiplies and distance between sites decreases, tending to concentrate in the most fertile areas and giving rise in some cases to cooperation strategies. We can infer, therefore, that the two periods analysed correspond to the thriving stages of Roman and Islamic political structures in our territory. After an initial phase of colonisation of the rural landscape, where settlement patterns tend to be almost random, farmers and peasants spread and the habitat stabilises in an aggregated way. Whereas the traditional Ripley's function cannot identify this process, our methodological approach detects the general underlying tendency towards clustering. In the social landscape created by the Roman city of Dertosa and later by the Islamic state of Ṭurṭūša, our data reflect a phase of prosperity and growth in the rural structures during peace times.

Case Study 2: Hunter-Gatherer Reoccupation Patterns in a Closed Bay

Study Problem and Materials

The Cambaceres Bay area (Beagle Channel, Tierra del Fuego, Argentina), with an extension of about 10 km^2, has been the focus of several archaeological studies on hunter-gatherer settlement and mobility strategies (Fig. 5). Those works included field surveys (Bjerck *et al*. 2016) and excavations (Orquera and Piana 1999; Álvarez *et al*. 2013) as well as specific analyses on palaeogeographic reconstructions (Zangrando *et al*. 2016), fishing resources (Zangrando 2009) and pinniped consumption patterns (Martinoli and Vázquez 2017), among other topics.

We centred our attention on the specific matter of hunter-gatherer mobility, which is thought to be related to aspects such as duration, functionality and reoccupation of a specific place (Kelly 1992; Bailey 2007; Zedeño and Anderson 2010; Negre *et al*. 2016a). In the particular case of indigenous societies settled on the Beagle Channel prior to the arrival of European expeditions, reoccupation patterns are thought to be a key aspect of mobility strategies. Archaeological methods have enabled the study of those patterns in the Fuegian region through the analysis of the disposition of their habitats. Ethnographic studies have contended that as long as a hut was still standing, reoccupation remained an option, increasing the density of the garbage deposits around it. These garbage mounds, also called shell middens due to the high proportion of marine molluscs in their composition, served both as basal shelter from climatic adversities and as a beacon for new settlements. When one of these accumulations reached an impractical size, the group tended to raise a new adjacent hut, developing over time an aggregated camp of non-contemporaneous occupations (Barceló *et al.* 2002a; Piana and Orquera 2010).

In this case study, we wanted to test the hypothesis that clustering exists among these hut camps. If positive, that might allow us to identify a clear pattern of reoccupation in the same areas, indicating that human settlement in one place favours subsequent reoccupation or occupations nearby over time. Our proposal is based on the data obtained during the archaeological surveys and test pit campaigns conducted between 2009 and 2013 within the framework of the Marine Ventures project (Bjerck *et al.* 2016), carried out by the Norwegian University of Science and Technology (Trondheim) and the Centro Austral de Investigaciones Científicas (Ushuaia). Field work identified a total of 1236 structures of both dwelling pits and domes, comprising at least 42 shell midden camps, each characterised by clusters of 10 or more structures thought to represent the reoccupation processes mentioned above (Fig. 6).

Despite some early- and middle-Holocene radiocarbon datings, related mainly to the excavated sites of Imiwaia and Binushmuka, most of the chronological sequence corresponds to the late Holocene (Bjerck and Zangrando 2013; Bjerck *et al.* 2016). As the topography of the bay has changed little over the last five millennia (Zangrando *et al.* 2016), the main goal of the proposed analysis is to detect trends in the areas with the highest concentrations of shell middens. For our *longue durée* analysis, we were interested more in overall trends in landscape use than in the precise moment in which they took place.

Results

The site distribution analysis presents, once again, very different results depending on the use of the Euclidean versus cost-distance approach. Applying the traditional

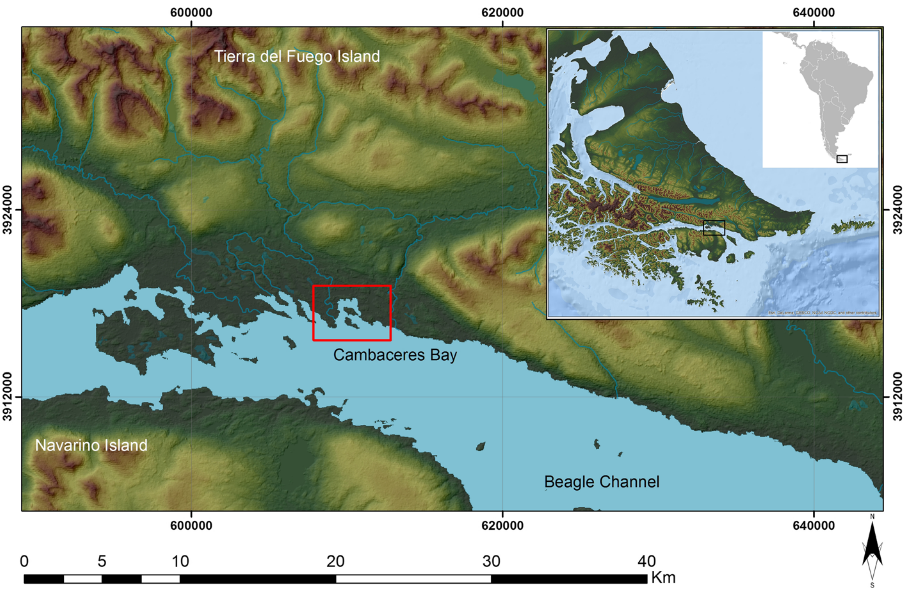


Fig. 5 Situation map

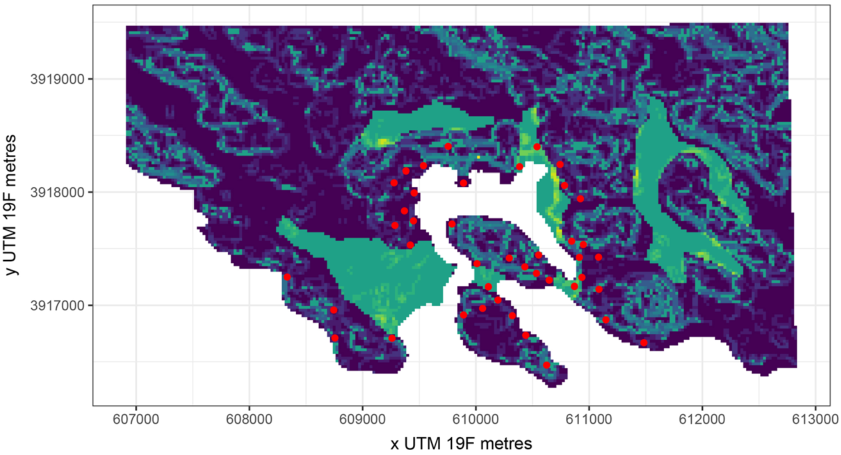


Fig. 6 Cost surface and sites along the Cambaceres Bay

Ripley's *K* function points to a spatial tendency within the margins of complete spatial randomness, showing a very slight tendency towards clustering starting at 450 m. On the other hand, the use of our cost-based approach reveals a significant pattern towards clustering at all scales. Furthermore, these results show that aggregations are very near to each other, suggesting higher-density clustering in more circumscribed areas (Fig. 7).

From a geographic viewpoint, the coast is an important source of heterogeneity in terms of spatial dependence. Euclidean-based approaches gave equal value to distances between sites placed on opposite sides of the bay as to those on the same side. However, these straight-line measures are misleading from a practical point of view given the restrictions on movement associated with crossing the frigid waters of the Beagle Channel—impossible without appropriate means or equipment.

Succinctly, shell mounds in this bay tend to be aggregated, thus forming large structured midden camps. Nevertheless, excavations suggest that this aspect does not necessarily account for social aggregation in the formation processes of the sites but sequences of individual depositions over time (Piana and Orquera 2010). The shore is an important feature where the largest structures are located, while the medium- and small-sized camps tend to occupy space at some distance, around the top of the low drumlin hills (Bjerck *et al*. 2016). When a nomad human population depends on marine resources for survival, vicinity to the shoreline and availability of firewood emerge as a

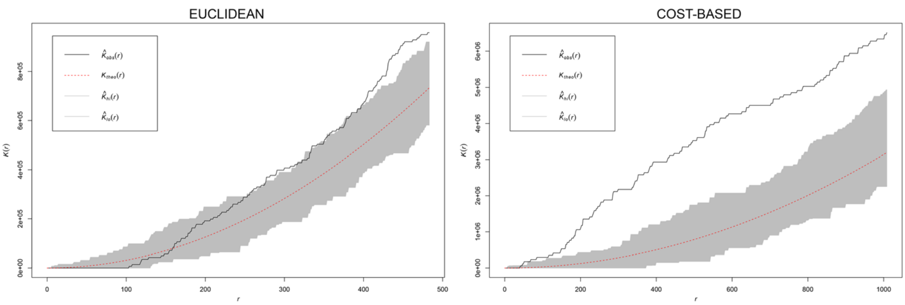


Fig. 7 Comparison between Euclidean and cost-based Ripley's *K* function

basic assumption. Therefore, the clustering pattern detected for the shell midden camps in this bay cannot be solely attributed to the proximity to these features, as the entire shoreline presents the same main characteristics. Space, in this case, is not a cause but a material product of social processes arising from social structures by determining future locations. Thus, human occupation begets reoccupation and occupation of neighbouring spaces, with evidence indicating that the generic presence of those camps was the direct cause of future settlements.

Comparing and Adjusting Approaches

The two-dimensional MDS (*cfr.* 2, 2.1.3) representation captures 75.7% of the variability in the cost-based distances for the Roman rural settlement patterns; it rises to 81.3% for the Islamic rural settlements and to 92.3% of the variability in Cambaceres Bay. In all cases, the cost-based distances are around an order of magnitude higher than the Euclidean distances. In a Euclidean space, the cost of crossing each cell of the raster is equal to 1 per metre, but the friction surface implies an exponential increase in that energetic expenditure. All three examples present a main positive gradient with linear association. The medieval phase shows the strongest positive correlation between the values, meaning that the results of the two approaches will differ less drastically. There are no obvious outliers in the visualisation results of the samples.

We also assessed whether any adjustments in the function are statistically necessary due to different effects. Border, translation and isotropic-corrected estimations for both Euclidean and cost-based approaches show no significant differences in any case study. Only the cost-based approach to the ancient Roman settlement pattern presents a slight decrease in clustering beginning at 35,000 cost units when edge-corrected (Fig. 8).

Discussion

We present two case studies, covering both macroscale and mesoscale territories, with significant variations in the results according to the choice of the Euclidean versus cost-based approach. The results of the latter show a highly convincing improvement in the accuracy and trustworthy of settlement pattern analysis. The higher the irregularity of the region under study, the more pronounced the divergence between the results generated by the traditional function and ours. Applying this method to a wide range of spatial processes and case studies should improve assessment. As we presented for microscale geostatistics, the use of non-Euclidean metrics for spatial correlation measurements is not a minor aspect to take into consideration (Negre *et al*. 2016b). Only by modelling the geographic source of spatial variability through the use of cost distances can our results be used as valid indicators of the underlying social processes in a settlement pattern.

Therefore, the most important contribution of this work lies in the possibility of finally overcoming the assumption of a homogeneous and undifferentiated surface of analysis. Actual human landscapes offer topographically dependent movement environments, and spatially heterogeneous natural resources and social attractors that cannot be overlooked in the historical analysis (Bevan and Conolly 2006). The

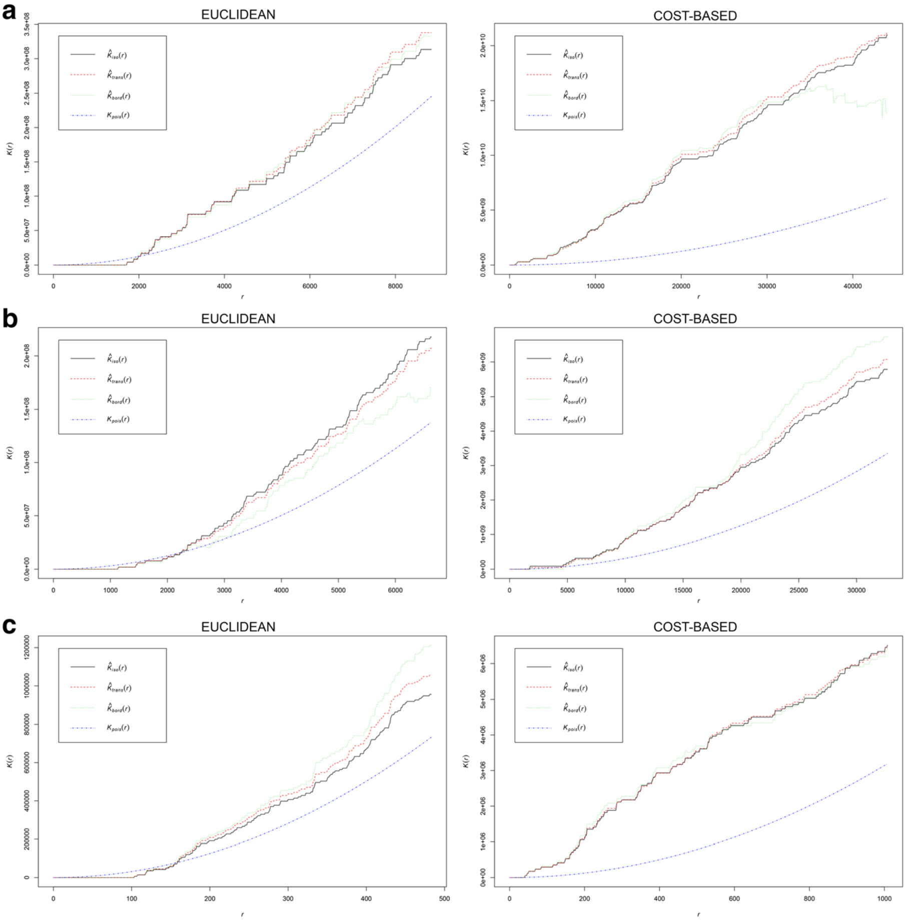


Fig. 8 Different corrections to the Ripley's *K* function for Euclidean and cost-based approaches for the Roman (**a**), Caliphate (**b**) and hunter-gatherer (**c**) settlements

Euclidean immersion presented in this paper could give more exact insight into the structure of archaeological spatial processes, taking into account the heterogeneous surface in which they took place. Within this new frame of reference, other spatial-statistical analyses (*e.g.* autocorrelation measures) could be performed, allowing a better understanding of the social phenomena under study.

Specifically, our paper addresses some of Bevan and Conolly's (2006) major concerns regarding the limitations in multi-scalar techniques applied to the study of settlement patterns: the reliability of point pattern analysis on Euclidean distance measures and the influence that site size or function might have in the overall process. Regarding the former, our approach overcomes the limitations imposed by current techniques. Moreover, thanks to the approximate Euclidean representation of the geographic pattern under study, weighted versions of Ripley's *K* function might be applied solving the latter aspect (Veen and Schoenberg 2006).

Finally, we want to emphasise a remaining concern regarding the study of settlement patterns: while technical improvements are increasing in quantity and quality, these advances are generally lacking in historical insight. Different mathematical models have demonstrated a number of significant relationships, both in studying settlement correlation with geophysical and social features and in analysing neighbouring patterns. Nevertheless, a settlement theory that generalises regularities in location strategies and spatial behaviours for different historical scenarios is still pending. In this line of research, the discussions on rural settlement and spatial behaviours available in the *Annals of the Association of American Geographers* are worth mentioning (Hudson 1969; Rushton 1969; Birch and Hudson 1970; Haining 1982). In those works, the authors proposed using spatial processes similar to those described for plant ecology in order to explain changes in habitat locations over time. Social dynamics like colonisation, spreading and competition, among others, were defined in terms of the distribution patterns they generated. Other works in this same direction defend, for example, the influence of urban cores (*i.e.* sites presenting a concentration of wealth and capital, individual enrichment and class formation, among other aspects) on the spatial structure of rural periphery (Barceló *et al*. 2002b; Macchi 2007; Negre 2013).

Our procedure may be useful for scholars interested in discovering how a particular spatial structure of distances between settlements can condition and determine the probability of social interaction, and how social and political factors (taxation, exchange, war, banditry, *etc*.) may have affected the particular location of residential and productive actions well beyond the limitations of topography. Social interaction is mostly a product of the social division of labour, so it is not a function only of observed spatial clustering, but of the differences and dependencies among social agents. This pattern is linked in turn to the flow of commodities (tools, raw materials, and manufactured goods), and information among social agents is linked through a network of spatial dependencies. Consequently, analysing the differences and dependencies among the location points of social actions should give us a better idea of how social interaction is built and reproduced, as well as the consequences of the specific means of interaction adopted.

The mere spatial distribution of archaeological sites cannot be used as a surrogate for identifying past social spaces. A point on a map signals where someone made something, used something produced elsewhere, and/or exchanged something from far away. These social actions were performed at particular places because of a social decision concerning the 'best' place, whether in terms of spatial efficiency, cost of mobility or some other consideration. Sometimes, this cost may have depended on physical factors (topography), but in many circumstances it would have depended on previous decisions regarding the probabilities of social and/or political interaction that transcend natural mobility (Barceló and Pallarés 1998). In this paper we have aimed to develop a theoretical framework by approaching the nature of spatial aggregation and social clustering as an obvious consequence of social interaction. In particular, we have examined the spatial relationship of different activities and whether such relationships sprang more from social decisions than the limitations imposed by the physical features of the land.

The development of models that capture the complex relationships between spatial regularity modalities and social processes can greatly complement ground-breaking methodological advances, allowing a better understanding of our reality. Interdisciplinary teams dealing with cross-cutting problems are currently the most important means to understand and process archaeological information.

Conclusions

Our proposed pseudo-Euclidean immersion, in which we can operate with conventional spatial-statistical tools, has proven to be a practical alternative for archaeological analysis of settlement patterns. The cost-based calculations produce more reliable results, avoiding the unrealistic assumption of a homogeneous study area. The reductionist application of the classical Ripley's function has been popular not because it fits the prevalent models of social dynamics but because it is the most straightforward way to work with currently available spatial statistical software packages. Nevertheless, our framework does not impose a conceptual model or a specific kind of interpretation.

In this paper, we have argued that a statistical index like Ripley's *K* function can be used to test whether places where a particular activity took place were clustered as a result of social decisions, independently of the natural advantages that such a place had from a topographical point of view. Our proposal is based on the use of a null hypothesis of lack of social intentionality instead of a null hypothesis of supposed spatial randomness. We have argued that this lack of human intervention does not take the form of a pattern of Euclidean equidistance between points—it should be analysed as a pattern of distances correlated to the cost surface of that particular landscape. In this way, we are enlarging the usual definition of social space, considering not only the physical features of a geographic area, but also the social uses of these features (rivers, mountains, plains, woods, *etc*.).

Space and time are the fundamental dimensions of change and dynamics, which have to be described, measured and understood. Spatial analysis involves operations whose results depend on data locations and thus cannot be reduced to the production of maps from a simple manipulation of the attribute database. In that way, we call for improving quality of spatial-statistical methods but, above all, relating those methods and their results to a different theories and scenarios regarding the social uses of space.

Acknowledgements We wish to thank Dr. Atilio F. Zangrando and Dr. Hein B. Bjerck who allowed us to use the data from Cambaceres Bay surveys in order to illustrate the benefits of our method. The present work benefitted from the input of Benjamin Ducke and Giancarlo Macchi who provided valuable comments to the undertaking of the method presented here. Joan Negre is part of the OCORDE Research Group, a recognised excellence group of the Generalitat de Catalunya [SGR2014-1454].

Funding Information

This work was supported by Spanish Ministry of Economy and Competitiveness through the project Genesis of the medieval nobility: fortresses and land control in northeastern Iberian Peninsula between the 8th and the 11th centuries [HAR2015-63661-P; MINECO-FEDER], directed by Dr. Ramon Martí at the Universitat Autònoma de Barcelona. Facundo Muñoz's research is partially supported by a research grant from the Spanish Ministry of Economy and Competitiveness (MTM2016-77501-P). Funding from Spanish ministerial projects Ethnicity in the Past [HAR2012-31036] and Computational simulation of sociohistorical and technological change phenomena during the Neolithic [HAR2016-76534-C2-1-R] are also acknowledged. We also thank Meggan Harris at Rosetta Manuscripts for her helpful comments and English language editing.

References

Álvarez, M., Fiore, D., Tivoli, A., Salvatelli, L., Saletta, M. J., & Briz, I. (2013). Variabilidad de actividades humanas en momentos recientes de la ocupación del canal Beagle (Tierra del Fuego): el caso de Lanashuaia XXI. In A. F. Zangrando, R. Barberena, A. Gil, G. Neme, M. Giardina, L. Luna, C.

Otaola, S. Paulides, L. Salgán, & A. Tivoli (Eds.), *Tendencias teórico-metodológicas y casos de estudio en la Arqueología de la Patagonia* (pp. 559–568). Altuna: Buenos Aires.
Baddeley, A., & Turner, R. (2015). Spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software, 12*(6), 1–42.
Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial point patterns: methodology and applications with R*. London: Chapman and Hall/CRC Press.
Bailey, G. (2007). Time perspectives, palimpsests and the archaeology of time. *Journal of Anthropological Archaeology, 26*, 198–223.
Barceló, J. A., & Pallarés, M. (1998). Beyond GIS: the archaeology of social spaces. *Archeologia e Calcolatori, 9*, 47–80.
Barceló, J. A., Piana, E. L., & Martinioni, D. (2002a). Archaeological spatial modelling: a case study from Beagle Channel (Argentina). In G. Burenhult & J. Arvidsson (Eds.), *Archaeological informatics: pushing the envelope* (pp. 351–360). Oxford: Archaeopress.
Barceló, J. A., Pelfer, G., & Mandolesi, A. (2002b). The origins of the city. From social theory to archaeological description. *Archeologia e Calcolatori, 13*, 41–63.
Bell, T., & Lock, G. R. (2000). Topographic and cultural influences on walking the ridgeway in later prehistoric times. In G. R. Lock (Ed.), *Beyond the map. Archaeology and spatial technologies* (pp. 85–100). Amsterdam: IOS Press.
Bevan, A., & Conolly, J. (2006). Multiscalar approaches to settlement pattern analysis. In G. R. Lock & B. Molyneaux (Eds.), *Confronting scale in archaeology: issues of theory and practice* (pp. 217–234). New York: Springer.
Birch, B., & Hudson, J. C. (1970). On a theory for rural settlement. *Annals of the Association of American Geographers, 60*(3), 610–614.
Bjerck, H. B., & Zangrando, A. F. (2013). Marine ventures: comparative perspectives on the dynamics of early human approaches to the seascapes of Tierra del Fuego and Norway. *The Journal of Island and Coastal Archaeology, 8*(1), 79–90.
Bjerck, H. B., Zangrando, A. F., Breivik, H. M., Piana, E., & Negre, J. (2016). *Marine ventures: the Cambaceres Surveys (2009, 2011–2013)*. Trondheim: Norwegian University of Science and Technology.
Christaller, W. (1966). *Central places in Southern Germany*. New Jersey: Prentice Hall.
Church, T., Brandon, R. J., & Burgett, G. R. (1999). GIS applications in archaeology: method in search of theory. In K. L. Wescott & R. J. Brandon (Eds.), *Practical applications of GIS for archaeologists. A predictive modelling toolkit* (pp. 135–155). Boca Ratón: CRC Press.
Conolly, J., & Lake, M. (2006). *Geographical information systems in archaeology*. Cambridge: Cambridge University Press.
Cox, T. F., & Cox, M. A. (2000). *Multidimensional Scaling*. Boca Ratón: Chapman & Hall/CRC Press.
Ducke, B., & Kroefges, P. (2008). From points to areas: constructing territories from archaeological site patterns using and enhanced Xtent model. In A. Posluschny, K. Lambers, & I. Herzog (Eds.), *Layers of perception. Proceedings of the 35th international conference on computer applications and quantitative methods in archaeology* (pp. 245–251). Bonn: Dr. Rudolf Habelt.
Fotheringham, A. S., & O'Kelly, M. E. (1989). *Spatial interaction models: formulations and applications*. London: Kluwer Academic.
Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2000). *Quantitative geography: perspectives on spatial data analysis*. London: Sage.
Gaffney, V., & Stančič, Z. (1991). *GIS approaches to regional analysis: a case study of the island of Hvar*. Oxford: Oxbow Books.
Haining, R. (1982). Describing and modeling rural settlement maps. *Annals of the Association of American Geographers, 72*(2), 211–223.
Herzog, I. (2014a). A review of case studies in archaeological least-cost analysis. *Archeologia e Calcolatori, 25*, 223–239.
Herzog, I. (2014b). Least-cost paths—some methodological issues. *Internet Archaeology, 36*. https://doi.org/10.11141/ia.36.5.
Holland, S. (2008). Non-metric Multidimensional Scaling (MDS). Technical Report, University of Georgia.
Hudson, J. C. (1969). A location theory for rural settlement. *Annals of the Association of American Geographers, 59*(2), 365–381.
Kelly, R. L. (1992). Mobility/Sedentism: Concepts, Archaeological Measures and Effects. *Annual Reviewers of Anthropology, 21*, 43-66.
Kvamme, K. L. (1997). Ranters corner: bringing the camps together, GIS and ED. *Archaeological Computing Newsletter, 47*, 1–5.

Llobera, M. (2000). Understanding movement: a pilot model toward the sociology of movement. In G. R. Lock (Ed.), *Beyond the map. Archaeology and spatial technologies* (pp. 65–84). Amsterdam: IOS Press.

Llobera, M., & Sluckin, T. J. (2007). Zigzagging: theoretical insights on climbing strategies. *Journal of Theoretical Biology, 249*, 206–217.

Lock, G. R., & Harris, T. (2000). Introduction: return to Ravello. In G. R. Lock (Ed.), *Beyond the map. Archaeology and spatial technologies* (pp. XIII–XXXV). Amsterdam: IOS Press.

López-Quílez, A., & Muñoz, F. (2009). Geostatistical computing of acoustic maps in the presence of barriers. *Mathematical and Computational Models, 50*(5–6), 929–938.

Macchi, G. (2007). *Geografia dell'incastellamento*. Firenze: All'Insegna del Giglio.

Macchi, G. (2009). *Spazio e misura. Introduzione ai metodi geografici-quantitativi applicati allo studio dei fenomeni sociali*. Siena: Edizioni dell'Università.

Marble, D. F. (1996). The human effort involved in movement over natural terrain: a working bibliography. Technical Report: Ohio State University, Department of Geography.

Martí, R., & Negre, J. (2015). Assentaments i espais agraris medievals al Baix Ebre i al Montsià: una anàlisi diacrònica. *Estudis d'Història Agraria, 27*, 67–89.

Martinoli, M. P., & Vázquez, M. (2017). Pinniped Capture and Processing: A Comparative Analysis from Beagle Channel (Tierra del Fuego, Argentina). In: M. Mondini, A. Muñoz, P. Fernández (Eds.), *Zooarchaeology in the Neotropics*. Cham: Springer.

Marwick, B. (2016). Computational reproducibility in archaeological research: basic principles and a case study of their implementation. *Journal of Archaeological Method and Theory, 24*(2), 424–450.

Matthews, J. A. (1981). *Quantitative and statistical approaches to geography*. Oxford: Pergamon.

Mayer, C. (2006). Making use of distances: estimating parameters of spatial processes. In M. W. Mehrer & K. L. Wescott (Eds.), *GIS and archaeological site location modeling* (pp. 137–153). Boca Ratón: CRC Press.

Møller, J., & Toftaker, H. (2012). Geometric anisotropic spatial point pattern analysis and cox processes. Research Report Series, Department of Mathematical Sciences, Aalborg University.

Muñoz, F. (2012). Geoestadística en regiones heterogéneas con distancia basada en el coste. PhD Thesis, Universitat de València.

Muñoz, F. (2015). geoRcb: An extension of package geoR that Works with Cost-based Distances. R package v.1.7–6. https://github.com/famuvie/geoRcb.

Muñoz, F., & Negre, J. (2017). cbK: Cost-based clustering analysis of human settlements. R package v.1.0. Zenodo. https://doi.org/10.5281/zenodo.250466.

Negre, J. (2013). De Dertosa a Turtusa. L'extrem oriental d'al-Tagr al-A'là en el context del procés d'islamització d'al-Andalus. PhD Thesis, Universitat Autònoma de Barcelona.

Negre, J. (2014). La cerámica altomedieval de Tortosa (siglos VII-X). Una primera clasificación y análisis interpretativo. *Arqueología y Territorio Medieval, 21*, 39–67.

Negre, J. (2015a). Non-Euclidean distances in point pattern analysis: anisotropic measures for the study of settlement networks in heterogeneous regions. In J. A. Barceló & I. Bogdanovic (Eds.), *Mathematics and archaeology* (pp. 369–382). Boca Ratón: CRC Press.

Negre, J. (2015b). Producciones cerámicas en el distrito de Turtusa entre la Antigüedad Tardía y el Mundo Islámico (siglos VI–XII). In M. J. Gonçalves & S. Gómez-Martínez (Eds.), *10th international congress on medieval pottery in the Mediterranean*. Silves: Cámara Municipal.

Negre, J., Álvarez, M., Pal, N., Bas, M., Briz, I., Lacrouts, A. & Lasa, A. (2016a). Spatial variability and occupation intensity in hunter-gatherer sites from the atlantic coast of Tierra del Fuego, Argentina. *Arqueología Iberoamericana, 32*, 37–51.

Negre, J., Muñoz, F., & Lancelotti, C. (2016b). Geostatistical modelling of chemical residues on archaeological floors in the presence of barriers. *Journal of Archaeological Science, 70*, 91–101.

Orquera, L. A., & Piana, E. L. (1999). *Arqueología de la región del canal Beagle (Tierra del Fuego, República Argentina)*. Buenos Aires: Sociedad Argentina de Antropología.

Philipps, C. (2004). GIS and landscape analysis, or the cart before the horse? *Internet Archaeology, 16*. https://doi.org/10.11141/ia.16.4.

Piana, E. L., & Orquera, L. A. (2010). Shell midden formation at the Beagle Channel (Tierra del Fuego, Argentine). In D. Calado, M. Baldia, & M. Boulanger (Eds.), *Monumental questions: prehistoric megaliths, mounds and enclosures* (pp. 263–271). Oxford: Archaeopress.

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

Rogerson, P. A. (2002). *Statistical methods for geography*. London: Sage.

Rushton, G. (1969). Analysis of spatial behavior by revealed space preference. *Annals of the Association of American Geographers, 59*(2), 391–400.

Schabenberger, O., & Gotway, C. (2005). *Statistical methods for spatial data analysis*. Boca Ratón: CRC Press.

Schwarz, K. R., & Mount, J. (2006). Integrating spatial statistics into archaeological data modeling. In M. W. Mehrer & K. L. Wescott (Eds.), *GIS and archaeological site location modeling* (pp. 154–174). Boca Ratón: CRC Press.

Soule, R. G., & Goldman, R. F. (1972). Terrain coefficients for energy cost prediction. *Journal of Applied Physiology, 32*(5), 706–708.

van Etten, J. (2015). gDistance: distances and routes on geographical grids. R Package, version 1.1–9. https://CRAN.R-project.org/package=gdistance.

van Leusen, M. (2002). Pattern to process: methodological investigations into the formation and interpretation of spatial patterns in archaeological landscapes. PhD Thesis: University of Groningen.

Veen, A., & Schoenberg, F. P. (2006). Assessing spatial point process models using weighted K-functions: analysis of California earthquakes. *Lecture Notes in Statistics, 185*, 293–306.

Verhagen, P., Kamermans, H., van Leusen, M., & Ducke, B. (2010). New developments in archaeological predictive modelling. In T. Bloemers, H. Kars, A. van der Valk, & M. Wijnen (Eds.), *The cultural landscape and heritage paradox* (pp. 431–444). Amsterdam: Amsterdam University Press.

Wheatley, D., & Gillings, M. (2002). *Spatial technology and archaeology. The archaeological applications of GIS*. London: Taylor & Francis.

White, D. A., & Surface-Evans, S. (2012). *Least cost analysis of social landscapes. Archaeological case studies*. Salt Lake City: The University of Utah Press.

Woodman, P. E., & Woodward, M. (2002). The use and abuse of statistical methods in archaeological site location modelling. In S. Poppy, D. Wheatley, & G. Earl (Eds.), *Contemporary themes in archaeological computing*. Oxford: Oxbow.

Zangrando, A. F. (2009). Is fishing intensification a direct route to hunter-gatherer complexity? A case study from the Beagle Channel region (Tierra del Fuego, southern South America). *World Archaeology, 41*(4), 589–608.

Zangrando, A. F., Ponce, J. F., Martinoli, M. P., Montes, A., Piana, E., & Vanella, F. (2016). Palaeogeographic changes drove prehistoric fishing practices in the Cambaceres Bay (Tierra del Fuego, Argentina) during the middle and late Holocene. *Environmental Archaeology*. https://doi.org/10.1080/14614103.2015.1130888.

Zedeño, M. N. & Anderson, D. (2010). Agency and politics in hunter-gatherer territory formation. *Revista de Arqueologia, 23*, 10-29.