GEOSTATISTICAL MODELLING OF ANTHROPIC ACTIVITY MARKERS IN THE PRESENCE OF BARRIERS.

Joan NEGRE PÉREZ (corresponding author)
Austral Centre for Scientific Research
National Research Council for Scientific and Technological Research
B. Houssay, 200 E-9410 Ushuaia (Tierra del Fuego, Argentina)
negreperez@gmail.com

Facundo MUÑOZ
UR0588 Unité Amélioration Génétique et Physiologie Forestières,
INRA Val de Loire Orléans, 2163 Avenue de la Pomme de Pin,
CS 40001 Ardon, Orléans Cedex 2, F-45075, France
facundo.munoz@orleans.inra.fr

ABSTRACT

KEYWORDS

HIGHLIGHTS

1. INTRODUCTION

The analysis of chemical soil composition in archaeological domestic floors is becoming increasingly considered as an important topic for historical research. Mapping the distribution of certain combinations of elements allows us to understand the activities that were developed in these areas. This assumption comes from the idea that different social actions of production, consumption or distribution are the cause of variations observed in the material consequences detected through fieldwork. In this case, the variability on chemical soil composition is thought to be a valuable marker in order to detect, identify and analyse different activities in domestic contexts (Rondelli et al. 2014, Salisbury 2013, Middleton et al. 2010). The model connecting the concentration of particular residues (proxies) with the specific activities inferred from different information sources (archaeological experimentation, ethnoarchaeological reasoning, etc.) is defined as an anthropic activity marker. Nonetheless, the reading of chemical differential concentrations in archaeological floors is not exempt of critical reflections about its limitations (Vyncke et al. 2011, Dore and López Varela 2010, Wells 2010, Terry et al. 2004).

Geostatistical methods are usually used to model the results of this kind of analyses and hence facilitate its interpretation. These techniques provide a set of statistical tools specifically designed for spatial problems, in which prediction is required over a region of interest where some observations have been taken. Predictions are based on an underlying statistical model that can take additional information into account as explanatory variables. In addition, the prediction error can be estimated based on propagation of uncertainty. One of the main limits of the use of geostatistical methods for that purpose is the assumption of a homogeneous, unrestricted space of analysis. This premise fail when we consider the spatial demarcations and topography that affect the distribution of our phenomena over the study region. The analysis of chemical residues in a domestic unit floors is a classic example of this kind of situation, in which the walls of the house affect to the diffusion of these elements.

In this work, we develop a methodology for overcoming this problem. We propose the use of cost-based distances to quantify the correlation between sampling locations. In this way, we present an interesting case study about the distribution of chemical elements in domestic floors within a controlled ethnographical context in Northern Gujarat (India) (Rondelli et al. 2014). This paper explores the relative spatial variability of these residues, taking into consideration spatial demarcations, to provide a method for the detection and interpretation of specific areas of activity. Our technique, therefore, might improve substantially the identification of both clustering patterns and different processes of floor maintenance and postdepositional dynamics considered as a background noise (Rondelli et al. 2014, Pecci et al. 2013, Barba 2007, Lloyd and Atkinson 2004).

2. ON THE USE OF NON-EUCLIDEAN DISTANCES IN GEOSTATISTICS

Geostatistics is a branch of statistics that encompasses the techniques that apply to geographical analysis. We owe its origins to the works of D.G. Krige (1951) and G. Matheron (1963) in the central decades of the 20th century. There are several applications of geostatistical methods in a very wide range of disciplines, with the common problem of modelling a Stochastic Process over a continuous spatial region from a partial group of observations. This process of interpolation is commonly assumed to be Gaussian, isotropic and intrinsically stationary (Cressie 1993). Geostatistical modelling is based on the principle of spatial dependence, this is near events are more related than distant ones. Nevertheless, what means near and how we calculate it?

Interpolation techniques assume that the correlation between the elements of a group of observations is a function of the Euclidean distance between them. In other words, stationarity is often accepted to mean that the spatial point process has constant intensity and uniform correlation depending only on the lag vector between pairs of points (Møller and Toftaker 2012). Considering the inherent irregularity of geographical terrain, either the presence of barriers or the difficulty to cross a region are presented as a major problem for this technical requirement.

Imagine two locations at a given (Euclidean) distance such that they are significatively correlated, because of underlying relevant factors affecting both of them. Now put a barrier between them that blocks or absorbs the effect of the underlying factors. This obviously pulls the correlation down. Therefore, when some kind of barriers exist, the correlation depends on something other than this kind of metrics, which therefore cannot account for correlation by itself.

Methodologically, the first step in geostatistical processing is to fit the data and its empirical semivariogram function to a known parametric model. There is a variety of methods for estimating this correlation. Our approach here is to use maximum likelihood methods that fits the mean function and the parameters of the semivariogram function. Once fitted, the main interest for analysing purposes is spatial prediction. Kriging assumes that the distance or direction between sample points reflects a spatial correlation that can be used to explain variation in the surface. This technique is one of the most used approaches to this problem, in which a weighted average of the sample values is used to generate the prediction. This is, sample points near the prediction's location are given larger weights than those far away. The general formula for the interpolator is formed as a weighted sum of the data:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i)$$

where $Z(s_i)$ is the measured value at the ith location, λ_i an unknown weight for the measured value at the ith location, s_0 the prediction location and N the number of measured values.

Kriging determines these weights based on the semivariogram function, calculating them according to the value of the semivariogram, which is a funct1ion of the Euclidean distance (López-Quílez and Muñoz 2009). That seems to incur into the above mentioned error of assuming the validity of the spatial homogeneity premise. Thus, in certain cases, alternative measurements to Euclidean metrics, such as cost-based, Riemannian or pseudo-Euclidean ones, represent the distance argument r of the semivariogram function more naturally.

2.1 COST-BASED DISTANCES

Alternative measures to Euclidean distances have been largely tested in several disciplines. A multidimensional scaled reconfiguration of the spatial distribution has proved to be very useful in some cases, allowing to create a pseudo-Euclidean framework on which the analysis can be performed (Løland and Høst 2003; Negre 2015). A fast Fourier Transform has also been checked for integrating moving-average functions that may be used to generate a large class of valid, flexible variogram models. This transform allows us both to compute the cross-variogram on a set of discrete lags and to interpolate the cross-variogram for any continuous lag (Ver Hoef et al. 2004). In this same direction, recent works also propose the use Riemannian metrics associated to cost-based distances and Banach algebra using Kuratowski immersion (Muñoz 2012: 118). For its relative simple implementation, the use of cost-based distances directly into the covariance matrix of the Kriging, has proved to be a practical and competitive option for our research topic.

From a methodological perspective, the main goal of cost-based distances is to define the least cost path to reach a known point from each cell location in the original raster dataset. The calculation algorithms present the length of the irregular vectors formed by a spatial distribution using the shortest weighted distance; this is the path with least accumulated cost. For allowing these calculations, first we need to create a cost surface, which purpose is to assign an impedance value to each cell of a raster layer, that is, the ease with which it can be crossed. Formally, the resulting cost-weighted model can be defined as an f function, which describes for each cell of our model a real, positive value representing the difficulty to go through them, that is, its cost-weighted density. Therefore, the cost-weighted movement dx to the point x is $f(x)dx$. From this function, the cost of any path in A can be calculated as the integration of every cell in the

model, which is gone through (Muñoz 2012: 55). Thus, the cost of a path α between points $s_1, s_2 \in A$ will be

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

These distances also maintain the same general properties than their metric counterparts (Waller and Gotway 2004: 321):

- non-negativity $(d(x,y) \geq 0)$
- symmetry $(d(x,y) = d(y,x))$
- triangle inequality $(d(x,z) \leq d(x,y) + d(y,z))$

That implies that these measurements could be used into geostatistical functions. In spite of that, the results might be, in some cases, statistically irrelevant. The more homogeneous is the surface under study, the less significant are the changes with respect to the use of Euclidean measures. To obtain mathematical validity, the results must check the positive definition of the covariance matrix of the observations. This condition requires that for any n number, set of locations $s_1, \dots, s_n$ and complex set of coefficients $\propto_1, \dots, \propto_n$, the R function verify the next relationship:

$$\sum_{i,1=1}^{n} \bar{\alpha_i}\, \alpha_j\, R(d\big(s_i, s_j\big))$$

where d represents the cost-weighted distance between their arguments (Muñoz 2012: 201). Ultimately, and provided verification of the above validation factors, a functional model can be described providing the best linear unbiased prediction.

The adaptation of geostatistical analysis computation to these metrics is present in three major stages: empirical variogram computation, variogram model parameter fitting and the actual kriging prediction. Apart from observation data and prediction locations needed for standard kriging, we also need two Cost-Based distance matrices previously computed. One holds the distances between observation points, a symmetric square matrix. The second matrix holds the distances between observation points and the prediction location(s), so it is an *n* (observations) x *m* (locations) sized matrix.

The empirical variogram is computed from the observation data only. It classifies pairs of observations into groups according to their distance, and then computes an estimator of the theoretical variogram value for that distance based on the differences between the observed values. In order to make a cost-based empirical variogram it is enough to make the initial classification based on the cost-based distance values given in the corresponding matrix, rather than calculating Euclidean distances. Note that this modification produces a different grouping of observation pairs. Therefore, variogram estimates will be different. The variogram model parameter fitting is also made based on observation data only. It is typically accomplished through maximum likelihood methods, trying out many possible combinations iteratively and keeping the best. This implies computation of the covariance matrix for each combination being tested. All we need is to make sure that the covariance matrix is computed based on the cost-based distances provided by our previously computed matrix. Finally, there is the kriging prediction. At this point, the covariance model is assumed to be known. Here again, we need to make sure that the covariance matrix of the observations is computed with the cost-based distances. In addition, the covariance between observation points and prediction locations are to be computed in order to make predictions. So this is when the second of the cost-based distance matrices is to be used.

2.2 METHODOLOGICAL OVERVIEW

2.2.1 Computing the cost-surface

The first operation is to encode the spatial heterogeneity of the working area into one *cost-surface*. This imply some modelling decisions and assumptions, which are not technical but scientific in nature. In our case, we assume that the soil is homogeneous except for the areas with solid sunken structures. These structures will completely interrupt the continuity of the area, blocking the dispersion of substances.

This conceptual model yields a cost-surface with a constant value of 1 everywhere, except over the structures where it takes an infinite value. In practice, any value larger than the diameter of the region will suffice. Alternatively, can be more practical to work in the inverted scale of a *conductivity* surface. In this case, the values would be simply 1 for regular conductivity and 0 for no conductivity, or infinite cost. Any of these alternative surfaces can be easily produced from a digital representation of the region with any GIS software, or with other spatially capable software like R (R Core Team, 2015).

In our case, we imported the ESRI shapefiles describing the geometry of the structures into spatial classes defined in the R package sp (Pebesma and Bivand, 2005). Then we used the function $rasterize()$ from the raster package (Hijmans, 2015) to produce a discretized surface with constant value 1 over the region of interest, and 0 over the solid structures. We used a resolution of 20 pixels/m.

2.2.2 Computing the cost-based distances

The cost/conductivity surface is the object representing our model of the region, and from where the distances between locations can be computed. Specifically, two matrices of cost-based distances are required: one $n \times n$ matrix with distances among the n observations, and one $m \times m$ matrix with the distances between each observation to each one of the m prediction locations.

We used the centroids of the conductivity raster cells as prediction locations to simplify mapping, although any set of prediction locations can be used. The computation of the distance matrices can be also be performed using a GIS software or right within R. López-Quílez and Muñoz (2009) use the first approach with the help of a specific script $v.costdist.mat$ (Muñoz 2015b) for GRASS GIS (GRASS Development Team, 2010). For this study, we used the R package $geoRcb$ (Muñoz 2015b) instead.

This package provides the function $distmatGen()$, which automatically computes the two cost-based matrices given the coordinates of the observations and the conductivity surface. Internally, it leverages the package $gdistance$ (van Etten, 2015) for efficient computation of least-cost paths, while taking care of all the technical details. Ultimately, the cost-based distance is computed using the well-known Dijkstra's algorithm for finding shortest paths between nodes in a network

2.2.3 Using cost-based geostatistical algorithms

The R package $geoRcb$ extends some functions from the $geoR$ package (Ribeiro and Diggle, 2015), in order to make them capable of working with cost-based distances. Specifically, the functions $variog$ and $likfit$ feature an additional argument $dists.mat$ which takes a symmetric matrix of distances between observation locations. These functions are respectively used to compute empirical variograms and fit variogram models. Finally, the alternative function $krige.conv$ performs the cost-based spatial prediction through conventional kriging by taking the required distance matrices as the additional arguments $dd.dists.mat$ and $dl.dists.mat$.

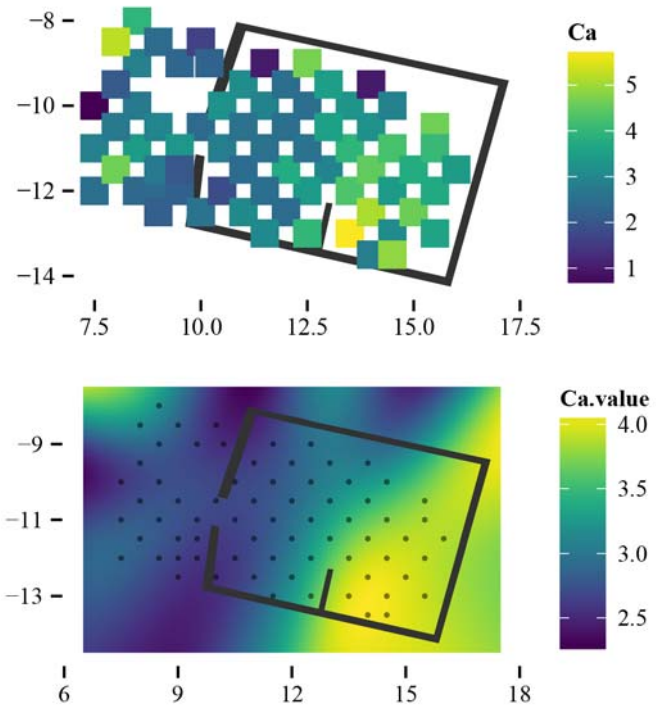
2.2.4 Presentation of results

We use the $viridis$ colour palette (Garnier 2015) for all the maps in the present paper. This palette is perceptually uniform and is also designed to be perceived by readers with the most common form of colour blindness.

3. CHEMICAL RESIDUES ANALYSIS ON ARCHAEOLOGICAL FLOORS IN THE PRESENCE OF BARRIERS

3.1 CONTEXT AND DATA DESCRIPTION

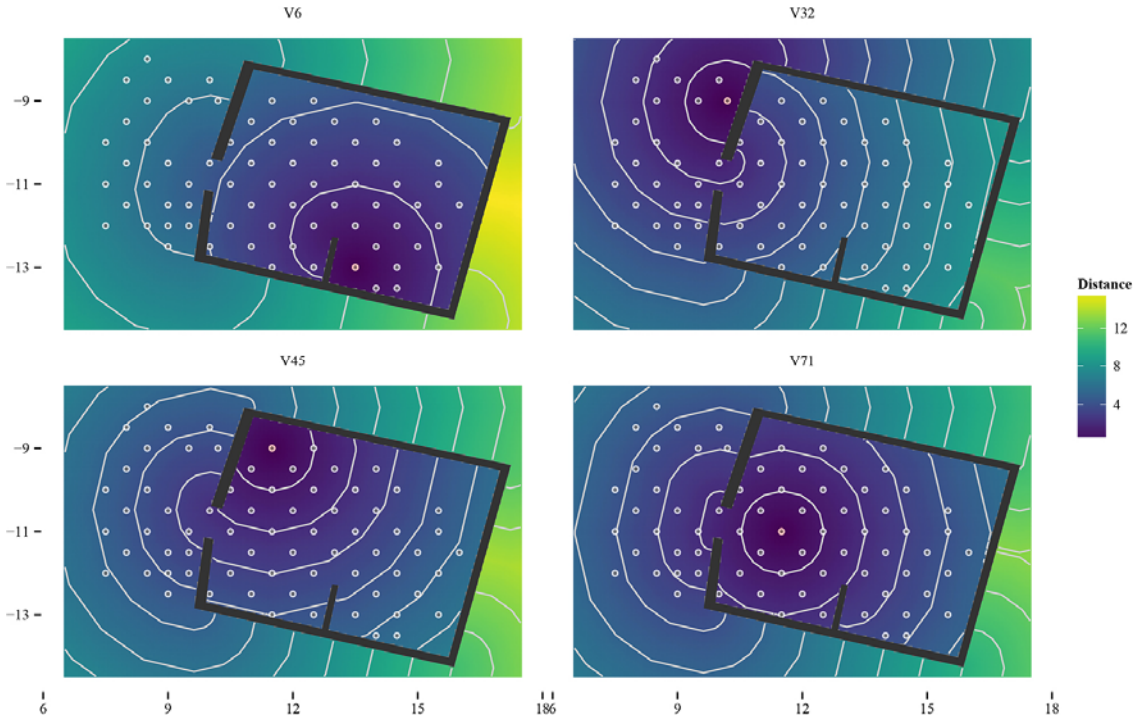
The data used in this methodological presentation was obtained from a fieldwork campaign carried out in a domestic compound in Jandhala, a village in the Patan district of North Gujarat, India. The compound is a non-mechanised farmer's residence built using local materials and traditional techniques. Several researchers of the North Gujarat Archaeological Project (NoGAP) participated on taking chemical soil samples in one of the domestic units of the compound between 2009 and 2010. A complete ethnoarchaeological study was developed in order to define the types and cycles of activities carried out in the compound, as well as their spatial distribution within the courtyard, veranda and inner house. Micromorphological floor samples were recovered during these works in order to obtain, among others, the chemical signatures of the average indicators of all the residues that fell on the floor since the foundation of the house, 10 years before sampling. Theses analyses were carried out with Inductively Coupled Plasma - Atomic Emission Spectrometry (ICP-AES), providing quantitative data on chemical elements. Finally, a complete study of the use of space based on floor chemical analysis was published (Rondelli et al. 2014). This is not the main goal of our work to analyse the specific use of space in this case study, but to confront the previous geostatistical results with the ones provided using our technique.



For testing purposes, we want to model a chemical spatial distribution map of Calcium residues in one of the domestic units of the Jandhala compound. Seventy measurements have been made in various points distributed homogeneously over the house and its courtyard and veranda. The Calcium values goes from a minimum of 0.66 to a maximum of 5.70, being its mean value 3.11 and its median 2.95. The data presents a normal or Gaussian distribution of the values, which stands for a non-random process (Barceló 2007: 45-49).

The walls of the house are considered non-transparent barriers for the diffusion of chemical residues on the archaeological floors. For this reason, cost-based distances between each sampling point to everywhere else have been calculated. Here, we show the cost-based distance

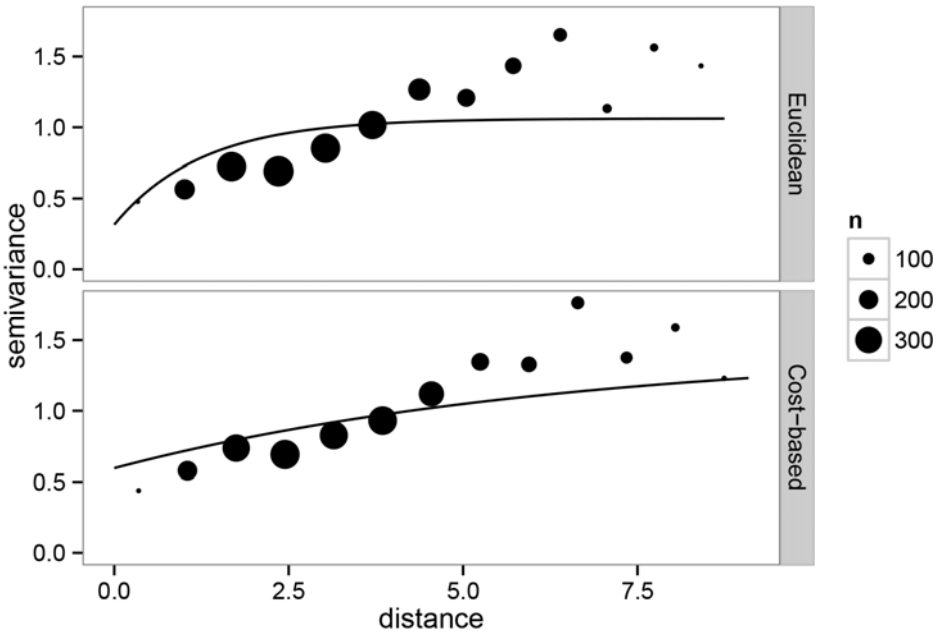
maps to four selected points, which present how these distances honour the geometry of the region:



3.2 VARIOGRAM AND FITTED MODEL

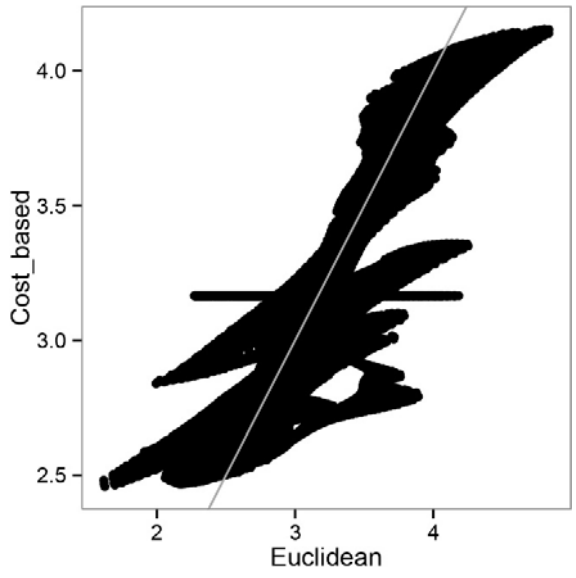
Variogram features (i.e., the estimated statistical parameters) include the nugget (the modeled discontinuity of the variogram at a distance of zero, which can represent measurement error or varia tion at distance too small to be captured with the current spatial design), the sill (the value at which the variogram stops increasing, an estimate of variance), and the range (the distance where the sill is reached, meaning the distance at which samples are no longer spatially correlated).

We fitted an exponential variogram model using both Euclidean and cost-based distances. The main difference being that the cost-based approach yields a larger range. This is usually the case, as incorporating the geometry into the analysis helps making sense of what was before interpreted as unstructured noise. Also the nugget is higher in the estimated cost-based variogram, which will result in increased smoothing. We choose to estimate the nugget effect, which may account for measurement error, for example. Near the observations, the cost-based approach has a larger prediction error due to its increased estimation of the nugget (i.e. short-range variance). In the main area, the prediction errors are practically the same with both approaches. Behind the walls, the Euclidean prediction error is unrealistically low. Leave-one-out Cross Validation (LOOCV) present similar error values for both approaches.



	Euclidean	*Cost_based*
Intercept	3.12	3.17
Nugget	0.32	0.60
Partial sill	0.75	0.85
phi	1.25	6.53
Pract. range	3.75	19.56

Once again, cost-based distance explains the correlation between locations better than Euclidean distance. A natural question to ask is how different the values from the two types of distances are. The following plot regarding pointwise comparison of predictions allows us to observe the high variability of the calculation depending on the method.

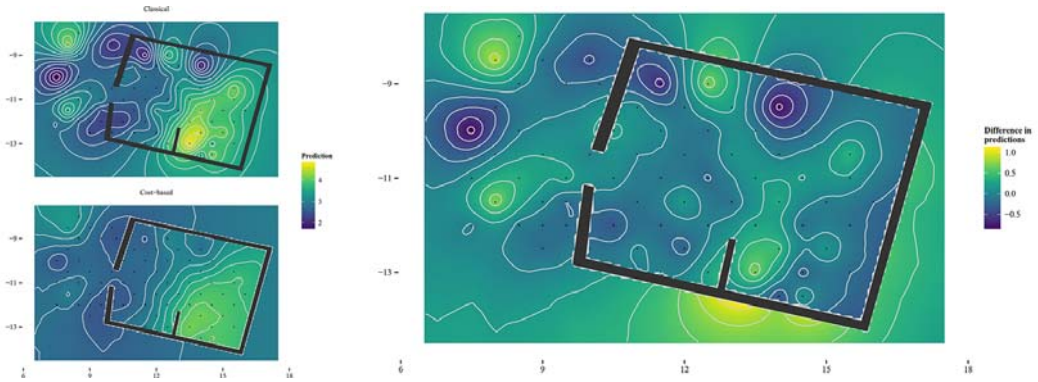


3.3 COMPARISON OF KRIGING ESTIMATES AND PREDICTED ERRORS

In the following image, it can be observed the map with the prediction for each location using Euclidean and Cost-based calculations under the same scale, and their difference. It is interesting to compare both results. There are three rmarkable aspects to be noted. First, the highest differences in prediction happen in the outer unsampled area, especially just across the wall from the most extreme observations. The Euclidean prediction is strongly influenced by these values, in contrast to the average value predicted by the cost-based approach. Second, the cost-based prediction is clearly smoother, as expected given the parameter estimates of the variogram model. This produces important differences in the neighbourhoods of the upper and lower measurements. Finally the walls of the house modify the prediction even in the sampled area. For example, in the corner near the central inner wall, across from the highest measurement, the Euclidean method predicts higher undoubtedly under the influence of the observation. On the contrary, in the veranda, the predictions are lower, influenced by the observations in the in the interior.

In summary, taking into account the heterogeneities of the working region using the cost-based distance leads to differences in predictions that can be split into these three main aspects. We believe that the cost-based approach produces more reliable results, avoiding the assumption of certain unrealistic premises about the homogeneity of the region.

Taking into account only the house and the veranda areas, cost-based approach is highly more accurate. The area of Calcium concentration is better defined through our approach, as well as the false positive areas, such as the external perimeter of the house or the veranda. The chemical representation of this element over the floor is perfectly demarcated close to the inner wall of the house, just in the south-east corner of it. That allows a better interpretation of the spatial distribution of this element and, especially, a better integration with other geochemical analyses. Archaeological reading of chemical signatures in domestic floors is highly related to the combination of diverse element residues, such as different combinations of Potassium, Calcium, Phosphorus, Magnesium and Strontium for identifying deposited food remains, living room sediments, enclosed spaces or burning areas (Rondelli et al. 2014: 488). Calcium has long been regarded as good indicator of human activity, usually correlated with Strontium, suggesting similar chemical pathways. Usually, these combined elements are interpreted as good indicators of covered, enclosed spaces. They should, then, serve as indicators of the integrity of anthropogenic deposits (Middleton et al. 1996: 679). In the presence of several geochemical indicators related at the same time, a perfect delimitation of their distribution areas result crucial for a valid inference. That plus of accuracy is what our approach provides.



This case study allows us to understand the main differences on the use of either non-Euclidean distances or their traditional metric counterpart. In this example, we are not focusing on the specific interpretation of Calcium hotspots in the Jandhala compound but in the different archaeological information obtained depending on the methodology that we used. This is an important question both in order to explain the distribution of this element and for using that information in combination with other proxies. The sum of small errors or imprecisions in the

process of modelling the distribution surfaces of each element is erroneously accumulated when those are combined in the final anthropic activity marker model. Regression Kriging using the combined trends that can be deduced from the statistical analysis of several chemical concentrations in each sampling point is also tied to this important limitation. Even so, spatio-temporal descriptions, such as interpolation maps, are not explanations but are themselves something to be explained (Barceló et al. 2015: 35). Therefore, when regression models are used to represent the general tendencies of geochemical residues, it is necessary to confront the interpolation of the resultant factor with the signatures of each individual sample for interpretation purposes.

4. DISCUSSION

Although the specific results of our case study, the proposed method is applicable in a general way toward interpolations in any archaeological intrasite example. As previously regarded, the more heterogeneous the surface under study (especially in the presence of barriers) the more needed this approach will be. The most important aspect of this work, therefore, lies in the general methodology for overcoming the geostatistical restriction on the homogeneity of the prediction region. In the study of geochemical signatures on archaeological floors, that limitation is a major problem in order to obtain significant interpretation of the sampling data. Thus, spatial distribution maps of chemical residues benefits from this methodology, since walls in domestic contexts are relevant restrictions in their diffusion. Furthermore, the possibility of applying geostatistical techniques enables us to obtain results based on statistical models, providing reliable predictions together with estimations of uncertainty, which commonly used deterministic methods cannot provide. Altogether, all this information allows the correct interpretation of the archaeological data, both the distribution of the chemicals and the postdepositional effects over them, linked among other factors, to the intensity of the uncertainty measure.

On the general objective of accurate readings on the spatial trends of chemical signatures, this proposal comes to increase a recent tendency on quantitative methods in our discipline. Since the ground-breaking work of Lloyd and Atkinson on the specific field of Geostatistics in Archaeology, a series of methodological milestones have taken place throughout the last decade. Our work wants to join efforts in developing this path by surpassing a recurrent limitation for these methods: the non-Euclidean characterisation of almost any geographical surface in the real world. As regarded for macro-scale archaeological cases this is not a minor aspect to take into consideration (Negre 2015). The treatment of heterogeneity and uncertainty, both in spatial and temporal spheres, has become the focus of most of the methodological step forwards to our discipline in the last years (Bevan et al. 2012; Deravignone et al. 2015; Fernández-López and Barton 2015; Crema 2015). The use of Anisotropic Geostatistics deals with both elements at the same time. The development of complex spatio-temporal reasoning in Archaeology is highly desirable when statistical, physical and chemical ground-breaking methodological advances are allowing a better understanding of our reality. Interdisciplinary work teams dealing with transversal problems are currently the most important path into new forms to understand and process archaeological information.

REFERENCES

Barba, L. 2007. Chemical residues in lime-plastered archaeological floors. Geoarchaeology 22: 439-452.

Barceló, J.A., 2007. Arqueología y Estadística. Introducción al estudio de la variabilidad de las evidencias arqueológicas. Universitat Autònoma de Barcelona, Bellaterra.

Barceló, J.A., K.F. Achino, I. Bogdanovic, G. Capuzzo, F. del Castillo, V. Moitinho and J. Negre, 2015. Measuring, Counting and Explaining: An Introduction to Mathematics in Archaeology. In J.A. Barceló & I. Bogdanovic (eds.). Mathematics and Archaeology (pp. 3-64). Boca Ratón: CRC Press.

Bevan, A., J. Conolly, C. Hennig, A. Johnston, A. Quercia, L. Spencer and J. Vroom, 2012. Measuring Chronological Uncertainty in Intensive Survey Finds. Archaeometry 55: 318-328.

Buck, C. and S.K. Sahu, 2000. Bayesian models of relative archaeological chronology building. Applied Statistics 49: 423-440.

Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley Series in Probability and Statistics. Wiley-Interscience, New York.

Deravignone, L., H.P. Blankholm and G. Pizziolo, 2015. Predictive Modeling and Artificial Neural Networks (ANN): from model to survey. . In J.A. Barceló & I. Bogdanovic (eds.). Mathematics and Archaeology (pp. 335-351). Boca Ratón: CRC Press.

Dore, C., S. López Varela, 2010. Kaleidoscopes, palimpsests, and clay: realities and complexities in human activities and soil chemical/residue analysis. Journal of Archaeological Method and Theory 17: 279-302.

Ebert, D., 2002. The potential of geostatistics in the analysis of fieldwalking data. In: Wheatley, D, G. Earl and S. Poppy (Eds.), Contemporary Themes in Archaeological Computing, University of Southampton Department of Archaeology Monograph No. 3. Oxbow books, Oxford, pp. 82-89.

Fernández-López de Pablo, J. and C.M. Barton, 2015. Bayesian Estimation Dating of Lithic Surface Collections. Journal of Archaeological Method and Theory 22: 559-583.

Garnier, S. 2015. viridis: Matplotlib Default Color Map. R package version 0.2. https://github.com/sjmgarnier/viridis

GRASS Development Team (2010). Geographic Resources Analysis Support System (GRASS GIS) Software. USA: Open Source Geospatial Foundation. http://grass.osgeo.org/

Hijmans R.J. (2015). raster: Geographic Data Analysis and Modeling. R package version 2.4-15. http://CRAN.R-project.org/package=raster

Krige, D.G., 1951. A statistical approach to some basi mine valuation problems on the witwaterstrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa 52(6): 119-139.

Lancelotti, C. and M. Madella, 2012. The "invisible" product: developing markers for identifying dung in archaeological contexts. Journal of Archaeological Science 39: 953-963.

Lloyd, C.D. and P.M. Atkinson, 2004. Archaeology and geostatistics, Journal of Archaeological Science 31: 151-165.

Løland, A. & Høst, G. (2003). Spatial covariance modelling in a complex coastal domain by multidimensional scaling. Environmetrics, 14 (3), 307-321.

López-Quílez A. and F. Muñoz (2009). Geostatistical computing of acoustic maps in the presence of barriers. Mathematical and Computer Modelling 50(5-6):929-938.

Matheron, G., 1963. Principles of geostatistics, Economic Geology 58(8): 1246-1266.

Middleton, W.D. and T.D. Price, 1996. Identification of Activity Areas by Multi-element Characterization of Sediments from Modern and Archaeological House Floors Using Inductively Coupled Plasma-atomic Emission Spectroscopy. Journal of Archaeological Science 23: 673-687.

Middleton, W.D., L. Barba, A. Pecci, J.H. Burton, A. Ortiz, L. Salvini and R.R. Suárez, 2010. The study of archaeological floors: methodological proposal for the analysis of anthropogenic residues by spot tests, ICP-OES, and GC-MS. Journal of Archaeological Method and Theory 17: 183-208.

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

Muñoz, F. M. (2012). Geoestadística en regiones heterogéneas con distancia basada en el coste. PhD Thesis: Universidad de Valencia.

Muñoz F. (2015a). geoRcb: An Extension of Package geoR that Works with Cost-Based Distances. R package version 1.7-5. https://github.com/famuvie/geoRcb. DOI:10.5281/zenodo.23568

Muñoz F. (2015b). v.costdist.mat: A GRASS script for computing cost-based distance matrices. Zenodo. DOI:10.5281/zenodo.23546

Negre, J. (2015). 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.

Pebesma E.J., R.S. Bivand, 2005. Classes and methods for spatial data in R. R News 5 (2), http://cran.r-project.org/doc/Rnews/

Pecci, A., M.A. Cau Ontiveros, C. Valdambrini and F. Inserra, 2013. Understanding residues of oil production: chemical analyses of floors in traditional mills. Journal of Archaeological Science 40 (2): 883-893.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ribeiro Jr P.J. and P.J. Diggle (2015). geoR: Analysis of Geostatistical Data. R package version 1.7-5.1. http://CRAN.R-project.org/package=geoR

Rondelli, B., C. Lancelotti, M. Madella, A. Pecci, A. Balbo, J. Ruiz Pérez, F. Inserra, Ch. Gadekar, M.A. Cau Ontiveros and P. Ajithprasad (2014). Anthropic activity markers and spatial variability: an ethnoarchaeological experiment in a domestic unit of Northern Gujarat (India), Journal of Archaeological Science 41: 482-492.

Salisbury, R.B., 2013. Interpolating geochemical patterning of activity zones at Late Neolithic and Early Copper Age settlements in eastern Hungary. Journal of Archaeological Science 40: 926-934.

Terry, R.E., F.G. Fernández, J.J. Parnell and T. Inomata, 2004. The story in the floors: chemical signatures of ancient and modern Maya activities at Aguateca, Guatemala. Journal of Archaeological Science 31: 1237-1250.

van Etten J. (2015). gdistance: Distances and Routes on Geographical Grids. R package version 1.1-7. http://CRAN.R-project.org/package=gdistance

van Overwalle, F. and D. van Rooy, 2013. A connectionist approach to causal attribution. In: S.J. Read and L.C. Miller (Eds.). Models of Social Reasoning and Social Behavior. (pp. 143-171). Routledge, New York.

Ver Hoef, J.M., N. Cressie and R.P. Barry, 2004. Flexible spatial models for kriging and cokriging using moving averages and the fast fourier transform (fft), Journal of Computational & Grafphical Statistics 13 (2): 265-282.

Vyncke, K., P. Degryse, E. Vassilieva and M. Waelkens, 2011. Identifying domestic functional areas. Chemical analysis of floor sediments at the Classical-Hellenistic settlement at Düzen Tepe (SW Turkey). Journal Archaeological Science 38: 2274-2292.

Waller, L. & Gotway, C. (2004). Applied Spatial Statistics for Public Health Data. Hoboken, Wiley.

Wells, E., 2010. Sampling design and inferential bias in archaeological soil chemistry. Journal of Archaeological Method and Theory 17: 209-230.

Wheatley, D. and M. Gillings, 2002. Spatial Technology and Archaeology: The Archaeological Applications of GIS. Taylor & Francis, London.