

Modelling Spatial Heterogeneity and Nonstationarity in Artifact-Rich Landscapes

Andrew Bevan and James Conolly

(Postprint of a 2009 paper in *Journal of Archaeological Science*
36.4: 956-964 (doi: 10.1016/j.jas.2008.11.023).

1 Introduction

Archaeological field surveys grapple with a palimpsested, partial, and heavily transformed record. Our efforts to explain and give meaning to this record are hampered by complex sequences of deposition, preservation and recovery, culminating in a particularly acute struggle against information degradation. Extensive (unsystematic) survey practices have their roots in the earliest stages of archaeology as a discipline, but it was really the development of more intensive and systematic techniques from the 1970s onwards—initially in several semi-arid and temperate parts of the world with frequent and highly visible surface artifacts—that illustrated survey’s potential to sample whole landscapes of human activity in ways that are amenable to statistical analysis (e.g., Flannery 1976; McDonald and Rapp 1972; Shennan 1985, to cite just three early important examples). However, despite three decades of important advances in data gathering and management techniques (e.g., Conolly and Lake 2006; Wheatley and Gillings 2002; Francovich and Patterson 2000, as recent examples) what

remains less well-developed are the quantitative methods for examining large multivariate, multi-scalar spatial datasets. Indeed, while there have been many recent advances in quantitative methods in geography, ecology and epidemiology (for a useful ecological review, see Miller et al. 2007), archaeology has been relatively slow to adopt and develop such approaches. It is telling that the last book-length treatment of spatial statistics in archaeology remains Hodder and Orton’s 1976 *Spatial Analysis in Archaeology*.

We improve on this situation by considering a crucial issue for survey archaeology: how, ultimately, do we identify and make sense of the heterogeneous and often inter-dependent behaviours and processes responsible for apparent spatial patterns? We argue that this issue is so fundamental to proper archaeological inference-building that it should not just reflect the methodological preoccupations of a statistically-enthusiastic minority, but can offer much wider theoretical capital. More precisely, we address two interconnected analytical problems—spatial heterogeneity and spatial nonstationarity—to illustrate how formal spatial modelling can offer useful insights into archaeological survey data and thereafter generate additional hypotheses about the cultural dynamics that created the archaeological record. To illustrate our claims, we draw on work produced by our Antikythera Survey Project (figure 1), where we have adopted explicitly interdisciplinary working practices, with GIS and spatial analysis prominent from the outset.

2 Problem definition

Between 2005 and 2007, we directed an intensive archaeological survey that sampled the entire landscape of the small ($\sim 20 \text{ km}^2$) Greek island of Antikythera via standard fieldwalking techniques. We also integrated a variety of geoarchaeological, botanical and ethno-historical studies in order to develop models of the

long-term human ecology and settlement history of the island. Our aims are not unique—there have been many surveys in the Mediterranean region that share similar objectives (e.g., Cherry et al. 1991; Jameson et al. 1994; Barton et al. 2004; Hill 2004; Barker et al. 2007, to name but five from a long list of possible examples). However, the geographic context of our study is more unusual; Antikythera is one of the smallest and most remote, yet frequently-inhabited, islands in the Mediterranean. It thus offers a rare kind of perspective on the various and variably-integrated ‘microregions’ that characterise the Mediterranean world (Horden and Purcell 2000: 124-143).¹

Our survey, to our knowledge, is the first to have surveyed an entire Mediterranean island with standardised field walking methods. First stage field walking was followed up by finer-scale second stage reinvestigation of particularly interesting areas: overall, we recorded approximately 100,000 artifacts and made a permanent collection of about a quarter of these, providing us with a highly structured and standardised archaeological dataset. In the course of analysis, it has been obvious, although not unexpected, that the artifact data exhibits multi-scalar patterning and, furthermore, that certain relationships between artifacts and environmental variables hold in some parts of the island, but not in others. This has prompted us to explore the basis of this observation in a more formal way and to investigate what contributing variables underpin it. The methods we used to achieve this are the focus of this paper; we believe they will potentially be of interest for a large number of archaeologists grappling with complex, artifact-rich, spatial datasets.

¹Our study program is also now complete and a final publication is in preparation. Preliminary datasets and results can be found at www.tuarc.trentu.ca/asp or www.ucl.ac.uk/asp.

3 The dataset

The raw data used for this analysis consists of counts of potsherds. These were recorded along transects walked by surveyors spaced 15 m apart and then standardised as estimated counts per 100 m², assuming for the moment that each surveyor consistently observes about 1 m of ground to either side of them as they walk (i.e., a “definite detection model”: Banning 2002). Rescaling of these densities for a different observed corridor would not affect the analysis or methods discussed here. Slightly more problematic is the likelihood that the real observation corridor from unit to unit varies depending both on the landscape covered (e.g. if it is treacherous and requires the surveyor to give less attention to surface observation) and the idiosyncrasies of the surveyor (and that detection rates also vary between observers; see Banning et al. 2006; Hawkins et al. 2003). However, for our practical purposes below, we have assumed a consistent observation corridor throughout. In Mediterranean survey, it is also common to aggregate individual walker observations into larger, polygonal units called ‘tracts’, which follow local field boundaries or vegetation patches and are often highly variable in size and shape (although typically sub-hectare). This practice can cause significant interpretative difficulties, as expressed by the well-known ‘modifiable areal unit problem’ (or ‘MAUP’, Fotheringham and Wong 1991; Amrhein 1995), which cautions that the type of patterns observable in aggregated datasets are to a greater or lesser extent dependent on both the scale and choice of aggregation unit (and particularly so for the measurement of correlations between variables). We have sought to reduce the impact of MAUP in our survey data by prioritising the original transect lines as our primary unit of recording and analysis, rather than tracts.

Surface pottery distributions are often skewed towards greater numbers of lower densities and this makes them difficult to represent with a straightforward,

linear colour scale (e.g. Dent 1999: 143–152). It also raises more fundamental analytical problems, not least because descriptive summary statistics such as the mean, standard deviation, etc., become far less useful and comparable across different landscape projects when the data do not have a normal, Gaussian, distribution. Of course, there are both exploratory and confirmatory statistical methods that are designed to work with skewed data, but many common techniques such as kriging and ordinary least squares regression are far less reliable under such conditions. A common treatment of skewed data thus is to normalise it by, for example, taking the logarithm (or if there are zero values present, adding one and then logging) and then later back-transforming the results if necessary. However, in many cases, artifact distributions resist even this sort of approach and arguably fall into an especially problematic type of data known as ‘zero-inflated distributions’ (see Lambert 1992 for the original discussion, and below for further references to spatial versions).

Second stage collections at finer resolution make it clear that in some cases, the abundance of zeros are false negatives arising from rapid observation time and disguise a range of low potsherd counts that would be identifiable if surveyors were to search for longer. From a theoretical standpoint, a clear starting point is to try to model these counts as one or more Poisson distributions (e.g. Orton 2008), albeit ones with (i) unusual rounding effects at the lower end and (ii) patterns of spatial and temporal autocorrelation (see below). Poisson distributions are often used to model the probability of a number of events occurring with a set temporal or spatial interval. They are also often used as an indication of complete spatial randomness within a particular distribution (Fotheringham et al. 2002b: 145). We will return to these issues at several stages in the discussion below.

4 Modelling spatial heterogeneity

In archaeology, spatial interpolation is widely used to model heterogeneous distributions of natural and cultural data for both visual and analytic purposes. The most common application of interpolation by archaeologists has traditionally been for the creation of digital elevation models (Conolly and Lake 2006: 100–107), but the technique can also be applied to model a continuous distribution of cultural data such as artifact densities or event horizons (e.g., Gkiasta et al. 2003) derived from sampled locations. The apparent complexity in archaeological spatial datasets arises from such spatial heterogeneity, insofar as the densities of different phenomena (e.g., artifacts, settlements, field walls, etc.) vary widely over space and, moreover, relationships that hold true in one area of a landscape may not hold true in another. To complicate things further, cultural (and natural) phenomena often also exhibit spatial anisotropy—i.e., measurements of a given phenomenon may not vary equally evenly in all directions, but may have trends that are directionally dependent (e.g., the frequency of artifacts may decrease evenly in one direction, but more chaotically in another). This complexity does not, however, mean that spatially heterogeneous and anisotropic phenomena are unable to be investigated quantitatively. In fact, as we here show, formal spatial analysis can yield a much deeper understanding of the structure of a dataset and its spatial variability.

4.1 Kriging

Kriging is a common technique in this regard, used extensively in geography and ecology, but less widely in archaeology (Zubrow and Harbaugh 1978; Robinson and Zubrow 1999; see also Conolly and Lake 2006: 97–101 and Lloyd and Atkinson 2004 for technical background). Kriging is a geostatistical technique that examines the spatial structure of a dataset to determine its interpolation param-

eters, and often provides more accurate and informative results than more basic interpolation methods. The essence of the technique is that instead of using a predefined weighting function that defines the contribution of each neighbouring sample point to the interpolated value, it is necessary to first model the degree to which the distance between observations is correlated (or not) to their values. This is established via the construction of a variogram, which provides information on the relative difference between observations, on average, when separated by a given distance (or ‘lag’). If this observed patterning can be accurately described by a theoretical model, then the modelling coefficients can be used to determine the interpolation parameters.

Kriging is thus more sensitive to the structure of the original data than interpolation functions that use an arbitrary distance or pre-defined number of neighbours to derive interpolated values (e.g., as in inverse-distance weighted interpolation). As we discussed above, the zero-inflated, Poisson character of our artifact count data makes traditional kriging analysis and variography problematic, but below we make use of a Poisson kriging approach which recent work in ecology and epidemiology has highlighted as a useful alternative under such conditions (Monestiez et al. 2006; Goovaerts 2005, Goovaerts 2008). An important additional component of kriging is establishing the degree of anisotropy in the dataset given that the type and intensity of directional relationships (e.g., whether information recovered from the north and south are more useful predictors than information recovered from the east or west) may provide additional insight into spatial structure of the phenomenon under examination.

The degree of anisotropy in a spatial dataset can be identified via variogram maps and/or directional variograms, both of which offer ways of visualising the decrease in spatial correlation at increasing scales of distance. Antikythera’s pottery counts suggest plenty of heterogeneous effects, but are more strongly

correlated with one another (i.e., showing lower semi-variance, γ) in a roughly NW-SE direction. In light of the larger scale heterogeneity, the directional variograms in figure 2b were calculated over a small spatial range where there is more consistency (up to ~ 400 m). Interestingly, both plots suggest the sampling effect of ASP’s survey lines, most of which were walked in either a N-S or E-W direction to facilitate easy mapping. This is possibly the source of the striped patterns in some parts of the variogram map, and as two groups (N-S, E-W versus NE-SW, SE-NW) of quite different semi-variances at the very shortest distances in the empirical variograms. Beyond these recovery effects, we can suggest that patterns of spatial autocorrelation are otherwise fairly isotropic over the first 100–150 m, but becomes far more anisotropic at larger distances with greater continuity in a NW-SE direction (i.e. over the ~ 400 m radius mentioned above, which we can describe more prosaically as covering areas 800 m or so wide).

After fitting a theoretical (spherical) model that describes this directional bias (Fotheringham et al. 2002a: 172), we can use the model’s coefficients to produce a prediction surface (the former shown in figure 3) for which we have much greater sense of the likely statistical error and a better understanding of the underlying spatial structure.² In any case, our interpretation of the variograms above is that their patterning arises from two separate processes: (i) the propensity for pottery counts on Antikythera to relate to one another at the level of small dense scatters (i.e. up to ~ 100 m across) and the emptier areas between them, and; (ii) at larger scales, for generally higher or lower densities to concatenate along the prevailing tectonic faultlines that split Antikythera up into valleys and hills running in similar NW-SE directions (and which encour-

²Although not explored here directly, the fitting of a separate variogram model for the very shortest-range variation might be a good way of controlling for the impact of the N-S or E-W striped, inter-observer variations discussed above (see Hawkins et al. 2003 for a good discussion of these effects and possible correction methods).

age the spatial dependence of a range of other phenomena as well, such as soils, slope, elevation, hydrology).

The statistical model of pottery density leads us to a range of more speculative, but testable hypotheses about artifact patterning, for example: (i) in one or more of the periods that are known to have produced the bulk of the surface ceramics on the island, there have been several clusters of activity of approximately 800 m in size, consisting of smaller sub-clusters, which are the products of larger, community-scale aggregations of small settlements; (ii) if the previous claim is correct and these 800 m size clusters do reflect community-level groupings, then each of the constituent sub-clusters should include similar types of functional activities, and have much in common in terms of the artifacts recovered; (iii) community-level groups will have particular spatial scales of correlation with other aspects of the environment such as soil, hydrology or topography, the distribution of which are also largely dictated by the major NW-SE geomorphological structure of the island.

5 Modelling spatial dependence and nonstationarity

Archaeological phenomena exhibit a strong degree of non-random structure both in time and space, and while this has long been seen as helpful for our taxonomic and chronological efforts (e.g., as in Childean space-time systematics), in many circumstances, it also can violate one of the key assumptions behind most ordinary statistical approaches, namely the assumption that individual observations are independent of one another. In fact such conditions are rarely present in the archaeological record. For example, the processes that produce surface artifacts are themselves often non-random—people produce large quantities of

pottery in certain locales (e.g., kiln sites), use sets of pottery in particular places (e.g., in houses), and often discard broken vessels together in dumps. This leads to high clustering of potsherds, and when we observe these in a series of archaeological survey units, the resulting sherd counts exhibit what is known as *spatial dependence*—broadly referring to situations where measured phenomena that are closer together in space tend to be more related than those further apart (popularly known as Tobler’s ‘First Law of Geography’—Tobler 1970). This property alone is enough to violate the assumptions behind many traditional statistical techniques, but is an explicit feature of the kriging analysis performed earlier. Spatially dependent patterns can have complex causes: some reflect the influence of exogenous factors, such as, in the pottery example, the correlation of high counts with land suitable for farming and settlement (itself a patchy, spatially-structured resource), while others indicate endogenous factors, such as the propensity for pottery to be discarded in groups and thereafter to break down from larger to smaller pieces (the proverbial ‘pot-smash’, though the latter issue might ideally be avoided by considering pottery weights rather than counts). As Fortin and Dale suggest, we can therefore usefully talk about *induced* spatial dependence (brought on by interaction with a range of exogenous influences) and *inherent* spatial dependence which reflects the self-influence of a particular property and which is better known as *spatial autocorrelation* (2005; see also Legendre 1993). A common example in ecology might be the spatial dependence in forest stands that results from varying geology on the one hand (an exogenous influence), and on the other, both seed dispersal mechanisms and subsequent competition between individual trees (endogenous cases). We will primarily focus on spatial patterns below, but it is worth noting that we can also talk about *temporal dependence*—where measured phenomena that are closer together in time tend to be more related than those further apart—and

distinguish between similarities that are induced by exogenous factors (e.g. the effect of climate change on cultural change) and those that reflect endogenous, autocorrelated effects (e.g. a classic case being patterns of cultural inheritance from one generation to the next).

These distinctions bring a greater level of analytical clarity but are insufficient on their own, because, we must also consider the possibility that the relationship between two or more variables across space may itself vary, which is termed *spatial nonstationarity*. As an example, consider a hypothetical relationship between two phenomena, x and y . A simple plot of one set of counts against the other might, for example, suggest a covarying relationship: where high quantities of x are found, so too are high quantities of y (and vice-versa). The problem is that, not only might the measured strength and statistical significance of this relationship vary spatially, but more dramatically, it might even be entirely different (e.g. negatively rather than positively correlated) in one part of the study area than in another. In both cases, such spatially varying relationships violate another key assumption behind most ordinary statistical approaches, namely the principle of *stationarity*. Nonstationarity effects mean that global correlation statistics such as Pearson's r^2 are likely to be inaccurate estimates of explanatory strength and that global models of the dependence of one spatial variable on another are likely, at the very least, to miss important local patterning, but also run the far more serious risk of mixing the effects of conflicting local relationships and producing an entirely spurious, mis-specified model (a manifestation of what is known as Simpson's Paradox). This can have profound implications for understanding the temporal and spatial relationship between cultural and ecological phenomena—what in fact we wish to understand is not simply *whether* x and y covary, but also *in what contexts, at what scales and in what local ways* do they covary. Traditional non-spatial statistics

are poorly-equipped to answer the latter questions, and we must thus look to advances in spatial statistics for help.

So far we have not made any attempt to model formally what exogenous environmental variables or endogenous processes might be inducing the spatial dependence of pottery counts on Antikythera, apart from suggesting that prevailing geology might play a role. In this respect, a more commonly adopted procedure is multiple regression. Traditional approaches to multiple regression derive a single set of relationships between one dependent variable and a number of independent variables. Perhaps the most common archaeological examples are the logistic regression models used to predict the relationship between the probability of encountering an archaeological site and several independent variables (such as distance to water, slope, elevation, etc.) for Cultural Resource Management purposes (for a well-constructed example, see Warren and Asch 2000). However, global models applied to geographic data are problematic for a variety of theoretical and methodological reasons, of which perhaps the most important but least recognised within archaeology is the fact that they assume a single set of stationary relationships between dependent and independent variables. In fact, this is extremely unlikely for many geographical phenomena, where non-stationary processes are common, or more prosaically, where “the measurement of a relationship depends in part on where the measurement is taken” (Fotheringham et al. 2002a: 9). In the real world, spatially varying relationships may be due to sampling bias (in the case of survey data, for example, potentially relating to the places explored by particularly observant or unobservant surveyors), but more interestingly, can also reflect the influence of a variety of environmental and cultural factors (e.g., spatially and temporally varying social attitudes, ecological niches, political structures, economic constraints, etc.).

Fortunately, there is now an increasing number of approaches that confront

and exploit the nonstationarity and spatial dependence present in multivariate relationships, which together are part of a general trend in the social sciences that has seen the decline of ‘general’ models in favour of ‘local’ ones (e.g., Lloyd 2007). These spatially-sensitive techniques are particularly common in ecology and epidemiology, including co-kriging and/or kriging with external drift, sparse grids, mixed autoregressive/trend models, and various Bayesian approaches (e.g., Goovaerts 2000; Cressie 1991; Green and Richardson 2002; Lichstein et al. 2002; Laffan et al. 2005). They are oddly rare, if not entirely absent, from the archaeological literature.

5.1 Geographically Weighted Regression

Here we introduce just one technique suited to the modelling of local patterning, known as geographically weighted regression (GWR, Fotheringham et al. 2002a). While a standard regression approach produces a single solution for (i) the intercept term, (ii) the coefficients that determine the weighting of independent variables, and (iii) the model’s goodness-of-fit, a geographically weighted regression estimates these separately at each sampled location. These estimates are applied locally within an area defined by a spatial kernel. A key decision therefore becomes the shape and size of this kernel (e.g. a circular neighbourhood of fixed radius or a consistent number of neighbouring sample points), as well as how the datapoints that fall within it will be weighted. Such a decision can be made by the user in a manner akin to the varying the focus on a (spatial) microscope, or can be optimised by iterative model-fitting and automated comparison of various goodness-of-fit statistics. The results can be compared to a global model to consider if there are any spatially varying relationships present and, if these are significant, the regression parameters and residual error can be mapped to explore the nature of these spatial patterns.

To illustrate, we first consider a traditional, global regression analysis and take the case of pottery counts on Antikythera as the example. As discussed in section earlier, we are dealing with count data that can be understood as the result of one or more Poisson processes and therefore modelled with reference to a Poisson distribution(s), but which show evidence of a greater number of zero counts than we might expect. An established way of dealing with excess zero counts is to perform a spatial, zero-inflated Poisson regression (‘spatial ZIP’—Agarwal et al. 2002; Rathbun and Fei 2006; Ver Hoef and Jansen 2007) which typically treats the phenomenon as a mixture of two components: (i) a binary distribution representing whether any counts were *detected* at a particular location, and (ii) a Poisson distribution which models the actual artifact counts. Conceptually, this approach might separate the detection (or not) of sherds from the measurement of the actual number of sherds present. As an aside, binary spatial data is amenable to formal statistical modelling (e.g., in a manner equivalent to that outlined by Rathbun and Fei 2006 for ecological data) that may help resolve the question of when zero counts in a survey actually reflect (i) an absence of surface artifacts versus (ii) the inability of surveyors to see artifacts that are present (e.g., because of vegetative interference), versus (iii) observer error. For example, spatial analysis of binary patterns may indicate where low frequency and isolated null counts are likely erroneous outliers, versus where they are clustered and spatially correlated with particular vegetation densities, which in turn may suggest the pattern reflects problems seeing surface artifacts.

However, the tendency for surface artifacts to cluster strongly together is itself a problematic factor as both this spatial clustering and the zero-inflation lead to count data which is ‘over-dispersed’ (i.e., with a variance larger than the mean) with respect to strict Poisson model assumptions. A second way of

addressing both of these problems, that is ultimately less conceptually satisfying than a ZIP model, but of much greater use for our purposes below, is to adjust the Poisson model using an estimated dispersion term (part of the quasi-family of generalised models; McCullagh and Nelder 1989: 323–371; Venables and Ripley 2002: 208–10; see also Bae et al. 2005). Although this works well here, we suggest that future studies should focus on how to manage zero-inflated distribution patterns in archaeological survey data by first considering the methodological advances that have been made by ecologists working with analogous datasets, and with particular attention to ZIP models.

Figure 4 and table 1 follow the simpler, second method suggested above and shows the setup and results of a quasi-Poisson, global regression when we include measures of slope, vegetation cover and geological variation as predictor variables. Slope is commonly known to influence not only settlement strategies and land use patterns, but also post-depositional factors such as erosion. Vegetation, as with slope, ultimately has a complex relationship with surface pottery distributions, but we can expect it to indicate more and less carefully managed parts of the modern landscape. For our limited purposes here, we calculated a Normalised Difference Vegetation Index (NDVI) from a QuickBird satellite image of Antikythera, which describes the amount of healthy green vegetation as rough proxy for changing land cover. High NDVI values represent dense vegetation, which correlates (albeit imperfectly) with areas of less intensive land management in the past and has important effects on recovery rates during archaeological survey. As another brief aside, many modern surveys include an assessment of percentage ground cover as part of their recording procedures and this serves primarily to indicate the amount of visual interference encountered when counting surface artifacts. It is then sometimes used to re-weight observed artifact counts (e.g., Given and Knapp 2003: 54-5; Bintliff et al. 2007:

21). However not only is the real relationship between ground cover and recovery rates extremely complex (see Schon 2000: 109; Bevan and Conolly 2004: 127-8; Given 2004: 16-17), but there are also fundamental problems associated with how consistently and accurately surveyors estimate ground surface cover. Our preferred solution has therefore been always to use raw artifact counts rather than weighted versions, and in cases where ground surface visibility or cover is of more general interpretative interest, to develop measures based on a high resolution QuickBird satellite imagery taken the year before the survey began (e.g. either the NDVI used here or a supervised classification of broad vegetation types).

Finally, we choose a continuous variable for geology which measures ‘distance into areas of harder limestone’. This allows us to model the influence of geology as a continuous field and thereby avoid the misleading edge effects (i.e. pot counts on either side of mapped geological boundaries) and awkward dummy regression variables that usually result from treating this parameter at the nominal scale. Antikythera consists of a combination of (i) older, harder limestones (Cretaceous-Palaeogene Gavrovo-Tripolitza unit) and (ii) softer, overlying flysch (probably Eocene), marls and conglomerates (Neogene; Galeos and Drandaki 1993; R. Siddall pers. com.) and this crude, binary difference determines the broad character of agricultural soils in different parts of the island, as well as the ease with which human structures (e.g. houses, walls, wells and tombs) could be made.

From the results in table 1, both vegetation and geology appear to be very significant predictors of pottery density on the surface of the island, with slope a more borderline case. Unfortunately, both visual mapping (figure 5f) and a Moran’s I test make it clear that the regression residuals (the remaining unaccounted-for variation) do not vary randomly across the landscape, but are

highly correlated. This suggests that the creation of one, global model is inappropriate because there are likely to be spatially-varying relationships amongst the predictors. GWR offers a method for addressing these nonstationarity issues. It is an approach that was first developed for linear regression, but can also be applied for any family of generalised linear model. Iterative comparison of the cross-validation scores for different kernel bandwidths suggests an optimal kernel with a fixed radius of about 420 m. Although the GWR kernel has a different form than the variogram model (see above and figure 2), they are strikingly close to one another in general size and suggest that beyond a threshold of about 800 m in diameter, spatially-dependent patterns become far more heterogeneous.

One of the most useful aspects of a GWR approach is the ability to map local estimates of the intercept, variable coefficients and other regression diagnostics and see how they vary across space. In figures 5a-f, most of the predictors seem to respond fairly consistently across the middle portion of the island, but elsewhere some are more spatially stationary than others. Slope, for example, has a slightly negative relationship with pottery density across much of the island, (i.e. as slopes get steeper, slightly fewer potsherds are found, figure 5b). However, this global relationship shows a much lower level of significance than the other two predictors (see table 1), and around the two harbours of Potamos and Xeropotamos in the north, the perceived global pattern potentially reverses itself, with a possible positive relationship (all other variables kept constant, and requiring that we confirm by testing for significance in this area alone). For NDVI, the strong negative correlation between increasing green vegetation (high NDVI values) and pottery seems most relevant to a limited number of areas (those in yellow), while much weaker relationships exist elsewhere and in the more barren areas of the far south and far east, there are again suggested

positive relationships. For geology, the overall pattern confirms the impression that human activity (as expressed by pottery discard) becomes less intense as you proceed further into zones of hard limestone, but the strength of this relationship varies and in certain key locales—the northern harbours, the far east and the far south—this relationship again breaks down. Such variation not only suggests that a global regression is highly misleading, but more positively, it encourages us to think carefully about how best to break the island into meaningful analytical regions. The dotted lines in figures 5a–d are a first effort to suggest some of these sub-regions, based on the varying regression relationships.

The sub-regions identified by GWR make some sense when we consider the contextual detail of settlement at local scales of analysis. To consider three of these localities: first, the central core of the island is arguably the most homogeneous part of the landscape and the one that is intensively used for farming in most periods. The stronger negative relationship between high slopes, NDVI values and limestone surfaces is thus not surprising and reflects a relatively consistent set of agriculturally-driven relationships in this area. Second, the far northern and far southern tips of the island are both fairly barren places and less agriculturally suitable, but while the former is bounded on all sides by cliffs and ridgelines, the latter has a small cove (with a modern lighthouse) and ships appear to have regularly harboured, resulting in elevated levels of pottery distributions, probably from the Bronze Age onwards. The determining relationships between pottery density and environment in this zone have thus not been adequately accounted for by our predictor variables, because they are linked to a range of other factors including coastal topography and morphology. Finally, the northern harbours require separate treatment not least because much of the intense activity that has occurred there, and the distinctively different relationships suggested by the regression analysis, reflect the impetus of

off-island connections. The most striking example of this is the area of extremely high pottery counts around one of the northern bays (the heavily over-dispersed distribution shown in red in figure 5e) which comes from the site of a Hellenistic ‘pirate’ community (described as such in at least one contemporary source: Jacopi 1932) perched in an otherwise unlikely location on a rocky headland, from which presumably to prey on the shipping lanes between Antikythera and Kythera to the north. Here too, the agriculturally-driven relationships between pottery and broad environmental characteristics identified for the central part of the island do not hold true, as the strategic priority for settling in the northern harbours area was very different.

GWR potentially therefore offers us a useful way to explore locally sensitive relationships between dependant and independent variables that are responsive to different sorts of land use patterns and settlement histories. However, the preceding analysis is meant primarily to demonstrate GWR’s promise as an exploratory technique, rather than as a definitive assessment of its formal analytical merits, and we would like to conclude by raising three concerns in this regard. First, there are methodological problems with the fact that GWR’s decomposition of the regression process into a series of small kernels can encourage much greater levels of multicollinearity between the variables (Wheeler and Tiefelsdorf 2005). In a sense, this is partly what it is designed to do—i.e., select a kernel size that maximises the amount of locally strong covariance—but consequently and more specifically, there is a very real risk of inducing greater levels of collinearity among the exogenous, predictor variables themselves and thereby making local parameter estimates inherently unstable. Second, although the residuals from this local regression are generally lower than for the global regression, they still show signs of autocorrelation that we would argue is often likely to be inherent rather than induced and present at smaller spatial scales

than the GWR kernel. A useful future approach to modelling this might be kriging of the residuals (e.g. Goovaerts 2000). Third and finally, while significance thresholds can be calculated for standard linear GWR models and should be presented alongside mapped parameter estimates (see Mennis 2006 for a guide to good cartographic practice in this regard), the task is problematic for more complex cases such as quasi-Poisson models and as suggested above, the parameter variations visible in figure 5 still require formal statistical confirmation within the proposed analytical sub-units before they can be wholly trusted. As always therefore, complex and relatively new statistical techniques need to be used with a degree of caution: in cases where an initial, global regression reveals only low-levels of spatial autocorrelation in its residuals, then this method (and for the above case, ideally a ZIP model) should be preferred. Where significant nonstationarity effects are present however, GWR should be seen as one of several promising ways of decomposing the study area into smaller, potentially more reliable units of analysis.

6 Conclusions

This paper has sought to address two under-appreciated issues associated with the spatial analysis of intensive survey datasets, but with the wider intention of developing more self-conscious and formal analysis of archaeological landscapes in general. Stratified sampling strategies have received a great deal of attention from landscape archaeologists (e.g., Barker et al. 2007 and Bintliff et al. 2007 as two applied examples; and Banning 2002 and Orton 2000 for theoretical models), and we would agree with much of this emphasis. In particular, and in the light of the issues raised in this paper, we would advocate multi-phase, multi-scale fieldwork methods: each strategic phase should involve collection units whose size, shape, spatial separation and observation time are as consistent as possible,

and where feasible, these should reflect some prior knowledge of the likely spatial scale and distribution of the archaeological phenomena involved (Banning 2002: 75). More tellingly, if intensive surface survey is to address patterns of spatial dependence effectively, it may be more advantageous to cover smaller study areas continuously rather than larger ones in many separate patches. However, it is not only our recovery strategies but also our analytical methods that must address the impact of spatially heterogeneous landscapes. Before we can consider more interesting relationships between cultural and environmental variables, we must seek methods that allow us to: (i) understand the scale at which a variety of spatial dependencies are operating, and (ii) make sense of spatially-varying relationships by breaking them down into simpler, more homogeneous units. Although not addressed directly here, such a perspective has relevance when we consider the temporal dependencies and temporal relationships present in archaeological datasets as well. In any case, such issues deserve greater attention than archaeologists have thus far given them.

Acknowledgements

A particular thanks to Aris Tsaravopoulos (our collaborator in the Greek Archaeological Service) for his assistance and encouragement. The Antikythera Survey Project was kindly funded by the Social Science and Humanities Research Council (Canada), the Institute for Aegean Prehistory and the Arts and Humanities Research Council (UK). We worked under a Greek Ministry of Culture permit and with the support of the Canadian Institute in Greece. The archaeological data used in this paper was managed and analysed in ArcGIS, GRASS, and R (including `spgwr`, Bivand and Yu 2007), as well as in a bespoke program for poisson kriging that was kindly supplied by Pierre Goovaerts. Our thanks also to Jeff Jenness for assistance in developing a semi-automated GIS

routine to plot walker lines within their larger tract polygons. An earlier version of this routine was first developed as part of the research conducted by both authors within a larger survey project on the neighbouring Greek island of Kythera since 1998 (the Kythera Island Project) and, more generally, we are very grateful to Cyprian Broodbank and Evangelia Kiriati (KIP co-directors) as well as the other KIP contributors for a wide range of advice, support and discussion. A final thank you to Clive Orton, Pierre Goovaerts and four anonymous *JAS* reviewers for their comments on the manuscript—their suggestions greatly improved the paper and our descriptions of spatial statistical methods. Any remaining issues or problems are, of course, our own responsibility.

References

- Agarwal, D. K., Gelfand, A. E., Citron-Pousty, S., 2002. Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics* 9, 341–355.
- Amrhein, C. G., 1995. Searching for the elusive aggregation effect: evidence from statistical simulations. *Environment and Planning A* 27 (1), 105–119.
- Bae, S., Famoye, F., Wulu, J. T., Bartolucci, A. A., Singh, K. P., 2005. A rich family of generalized poisson regression models with applications. *Mathematics and Computers in Simulation* 69, 4–11.
- Banning, E., Hawkins, A., Stewart, S., 2006. Detection functions for archaeological survey. *American Antiquity* 71 (4), 723–740.
- Banning, E. B., 2002. *Archaeological Survey*. Kluwer, New York.
- Barker, G., Gilbertson, D., Mattingly, D. (Eds.), 2007. *Archaeology and Desertification: The Wadi Faynan Landscape Survey, Southern Jordan*. Wadi

- Faynan Series 2, Levant Supplementary Series 6. Council for British Research in the Levant, Oxbow Books, Oxford.
- Barton, C. M., Auban, J. B., Puchol, O. G., Schmich, S., Balaguer, L. M., 2004. Long-term socioecology and contingent landscapes. *Journal of Archaeological Method and Theory* 11, 253–395.
- Bevan, A., Conolly, J., 2004. Gis, archaeological survey and landscape archaeology on the island of kythera, greece. *Journal of Field Archaeology* 29, 123–138.
- Bintliff, J., Howard, P., Snodgrass, A. (Eds.), 2007. *Testing the Hinterland: The Work of the Boeotia Survey (1989–1991) in the Southern Approaches to the City of Thespiiai*. McDonald Institute Monographs. McDonald Institute for Archaeological Research.
- Bivand, R., Yu, D., 2007. *The spgwr Package*. R Foundation for Statistical Computing.
- Cherry, J., Davis, J., Mantzourani, E., 1991. *Landscape Archaeology as Long-term History : Northern Keos in the Cycladic Islands from Earliest Settlement until Modern Times*. UCLA Institute of Archaeology, Los Angeles.
- Conolly, J., Lake, M., 2006. *Geographical Information Systems in Archaeology*. Cambridge University Press, Cambridge.
- Cressie, N. A. C., 1991. *Statistics for Spatial Data*. John Wiley, New York.
- Dent, B. D., 1999. *Cartography. Thematic Map Design*. McGraw-Hill, Boston.
- Flannery, K., 1976. *The Early Mesoamerican Village*. Academic Press, New York.
- Fortin, M.-J., Dale, M. R. T., 2005. *Spatial Analysis: A Guide for Ecologists*. Cambridge University Press, Cambridge.

- Fotheringham, A., Wong, D. W. S., 1991. The modifiable area unit problem in multivariate statistical analysis. *Environment and Planning A* 23, 1025–44.
- Fotheringham, A. S., Brunston, C., Charlton, M., 2002a. *Geographically Weighted Regression*. Wiley, Chichester.
- Fotheringham, A. S., Brunston, C., Charlton, M., 2002b. *Quantitative Geography: Perspectives on Spatial Data Analysis*. Sage Publications, London.
- Francovich, R., Patterson, H. (Eds.), 2000. *Extracting Meaning from Ploughsoil Assemblages*. No. 5 in *The Archaeology of Mediterranean Landscapes*. Oxbow Books, Oxford.
- Galeos, A., Drandaki, E., 1993. Geological map of greece. antikythera island sheet, (1:50,000). Tech. rep., Institute of Geology and Mineral Exploration., Athens.
- Given, M., 2004. Mapping and manuring: can we compare sherd density figures. In: Alcock, S. E., Cherry, J. F. (Eds.), *Side-by-Side Survey. Comparative Regional Studies in the Mediterranean World*. Oxbow Books, Oxford, pp. 13–21.
- Given, M., Knapp, A. B., 2003. *The Sydney Cyprus Survey Project: Social Approaches to Regional Archaeological Survey*. Cotsen Institute of Archaeology, Los Angeles.
- Gkiasta, M., Russell, T., Shennan, S., Steele, J., 2003. Neolithic transition in europe: the radiocarbon record revisited. *Antiquity* 77, 45–62.
- Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology* 228, 113–129.

- Goovaerts, P., 2005. Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using poisson kriging. *International Journal of Health Geographics* 4 (31), doi:10.1186/1476-072X-4-31.
- Goovaerts, P., 2008. Kriging and semivariogram deconvolution in the presence of irregular geographical units. *Mathematical Geosciences* 40, 101–128.
- Green, P. J., Richardson, S., 2002. Hidden markov models and disease mapping. *Journal of the American Statistical Association* 97 (460), 1055–1070.
- Hawkins, A. L., Stewart, S. T., Banning, E. B., 2003. Interobserver bias in enumerated data from archaeological survey. *Journal of Archaeological Science* 30, 1503–1512.
- Hill, J. B., 2004. Time, scale and interpretation: 10,000 years of land use on the transjordan plateau, amid multiple contexts of change. In: Athanassopoulos, E. F., Wandsnider, L. (Eds.), *Mediterranean Archaeological Landscapes: Current Issues*. University of Pennsylvania Museum, Philadelphia, pp. 125–142.
- Hodder, I., Orton, C., 1976. *Spatial Analysis in Archaeology*. New Studies in Archaeology. Cambridge University Press, Cambridge.
- Horden, P., Purcell, N., 2000. *The Corrupting Sea*. Blackwell, London.
- Jacopi, G., 1932. Nuove epigraphi dalle sporadi meridionali. *Clara Rhodos* 2, 165–256.
- Jameson, M. H., Runnels, C. N., van Andel, T. H., 1994. *A Greek Countryside*. Stanford University Press, Stanford.
- Laffan, S. W., Nielsen, O. M., Silcock, H., Hegland, M., 2005. Sparse grids: a new predictive modelling method for the analysis of geographic data. *International Journal of Geographical Information Science* 19 (3), 267–292.

- Lambert, D., 1992. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics* 34 (1), 1–14.
- Legendre, P., 1993. Spatial autocorrelation: Trouble or new paradigm? *Ecology* 74 (6), 1659–1673.
- Lichstein, J. W., Simons, T. R., Shriver, S. A., Franzreb, K. E., 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs* 72 (3), 445–463.
- Lloyd, C. D., 2007. *Local Models for Spatial Analysis*. CRC/Taylor and Francis, Boca Raton.
- Lloyd, C. D., Atkinson, P. M., 2004. Archaeology and geostatistics. *Journal of Archaeological Science* 31, 151–165.
- McCullagh, P., Nelder, J. A., 1989. *Generalized Linear Models*. Chapman and Hall, London.
- McDonald, W., Rapp, G. (Eds.), 1972. *The Minnesota Messenia Expedition: Reconstructing a Bronze Age Regional Environment*. University of Minnesota Press.
- Mennis, C. J., 2006. Mapping the results of geographically weighted regression. *The Cartographic Journal* 43 (2), 171–179.
- Miller, J., Franklin, J., Aspinall, R., 2007. Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling* 202, 225–242.
- Monestiez, P., Dubroca, L., Bonnin, E., Durbec, J. P., Guinet, C., 2006. Geostatistical modelling of spatial distribution of balaenoptera physalus in the northwestern mediterranean sea from sparse count data and heterogeneous observation efforts. *Ecological Modelling* 193, 615–628.

- Orton, C., 2000. *Sampling in Archaeology*. Cambridge Manuals in Archaeology. Cambridge University Press, Cambridge.
- Orton, C., 2008. Horse kicks, flying bombs and potsherds: statistical theory contributes to archaeological survey. *Archaeology International* 10, 24–27.
- Rathbun, S. L., Fei, S., 2006. A spatial zero-inflated poisson regression model for oak regeneration. *Environmental and Ecological Statistics* 13, 409–426.
- Robinson, J. M., Zubrow, E., 1999. Between spaces: Interpolation in archaeology. In: Gillings, M., Mattingly, D., van Dalen, J. (Eds.), *Geographical Information Systems and Landscape Archaeology*. Vol. 3 of *The Archaeology of Mediterranean Landscapes*. Oxbow Books, Oxford, pp. 65–84.
- Schon, R., 2000. On a site and out of site: where has our data gone? *Journal of Mediterranean Archaeology* 13, 107–111.
- Shennan, S., 1985. *Experiments in the Collection and Analysis of Archaeological Survey Data: The East Hampshire Survey*. Academic Press, New York.
- Tobler, W. R., 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 42 (2), 234–240.
- Venables, W. N., Ripley, B. D., 2002. *Modern Applied Statistics with S*. Springer, New York.
- Ver Hoef, J. M., Jansen, J. K., 2007. Space–time zero-inflated count models of harbor seals. *Environmetrics* 18, 697–712.
- Warren, R. E., Asch, D. L., 2000. A predictive model of archaeological site location in the eastern prairie peninsula. In: Wescott, K. L., Brandon, R. J. (Eds.), *Practical Applications of GIS for Archaeologists: A Predictive Modelling Kit*. Taylor and Francis, London, pp. 3–32.

- Wheatley, D., Gillings, M., 2002. *Spatial Technology and Archaeology. The Archaeological Applications of GIS*. Taylor & Francis, London.
- Wheeler, D., Tiefelsdorf, M., 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems* 7, 161–187.
- Zubrow, E., Harbaugh, J., 1978. Archaeological prospecting: kriging and simulation. In: Hodder, I. (Ed.), *Simulation Studies in Archaeology*. Cambridge University Press, Cambridge, pp. 109–122.

Table

<i>Parameter</i>	<i>Estimate</i>	<i>Standard Error</i>	<i>p-value</i>
Intercept	2.55	0.14	< 2e-16
Slope	-0.01	0.01	0.06
NDVI	-4.92	0.57	< 2e-16
Geology	-0.002	0.0002	< 2e-16
<i>Estimated dispersion</i>	98.7		
<i>Null deviance</i>	192717 (12022 d.f.)		
<i>Residual deviance</i>	164784 (12019 d.f.)		
<i>Cross-validation score</i>	303.6		
<i>Moran's I test (residuals)</i>	187.6 (p < 2.2e-16)		

Table 1: Summary results from a global regression with Slope, NDVI and Geology as independent variables.

Figures

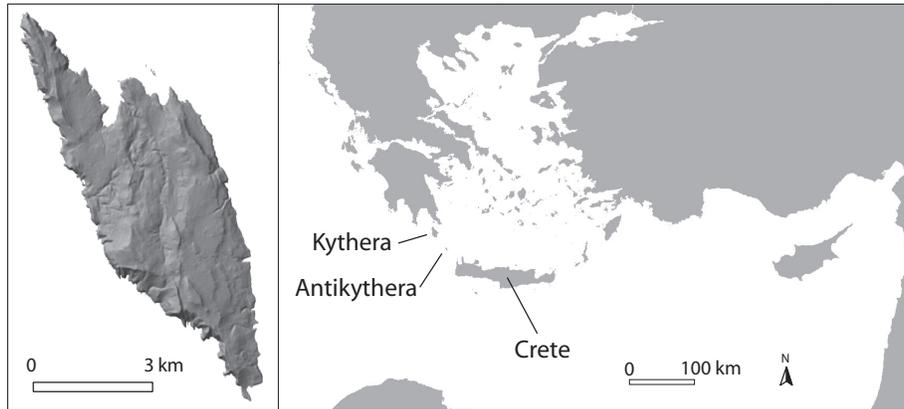


Figure 1: The south-western Aegean island of Antikythera (ASP intensively surveyed 19 km² of the island's total 20 km² extent), with a location map of the eastern Mediterranean.

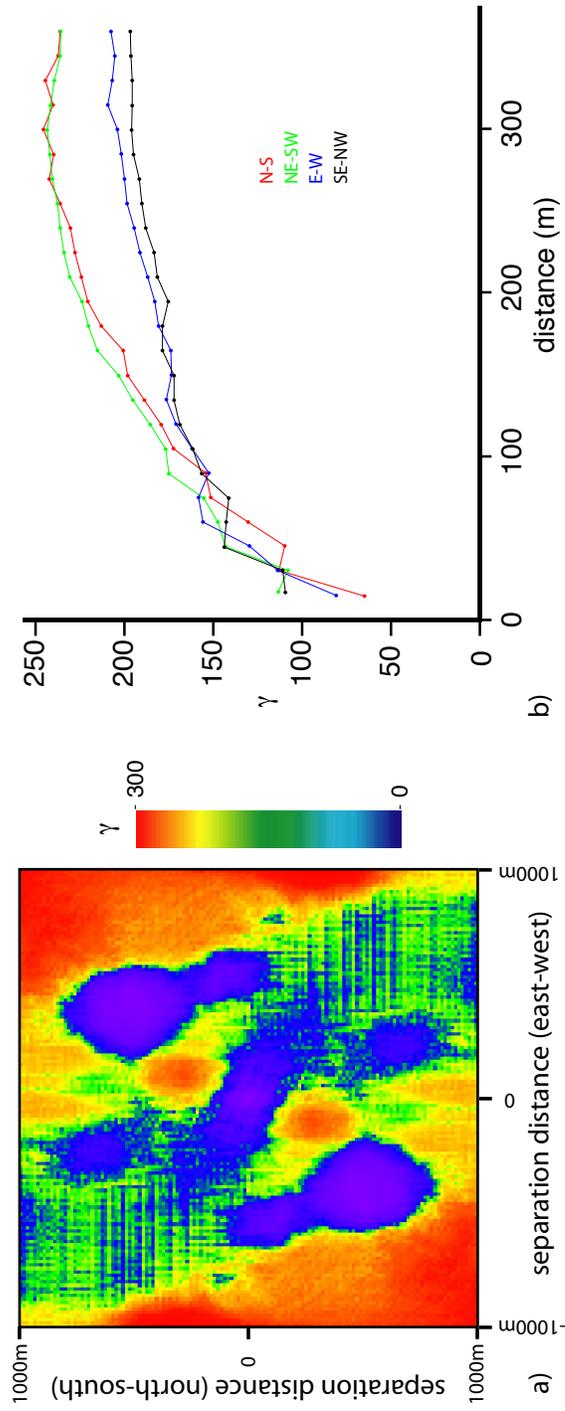


Figure 2: Anisotropic spatial dependence in the surface pottery counts from Antikythera: (i) a variogram map, and (ii) directional variograms (both with 15 m lag spacing).

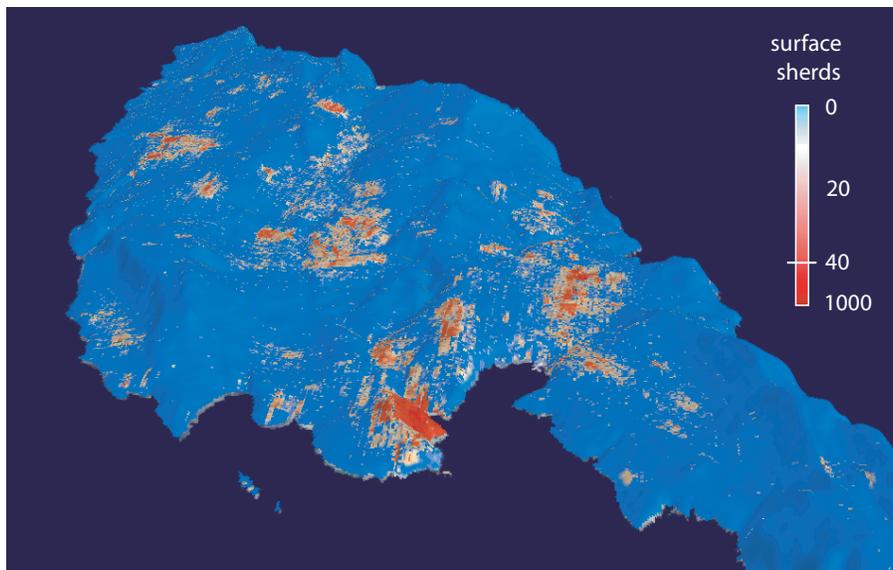


Figure 3: A visualisation of ASP surface pottery density, drawing on walker-scale records, looking south. Pottery densities include original estimated densities in areas densely covered by walker lines and, in-between, use estimates from poisson kriging with an anisotropic kernel. Results are draped over a digital elevation model and densities are expressed as sherd counts per 100 m² (i.e. per 10×10 m raster cell).

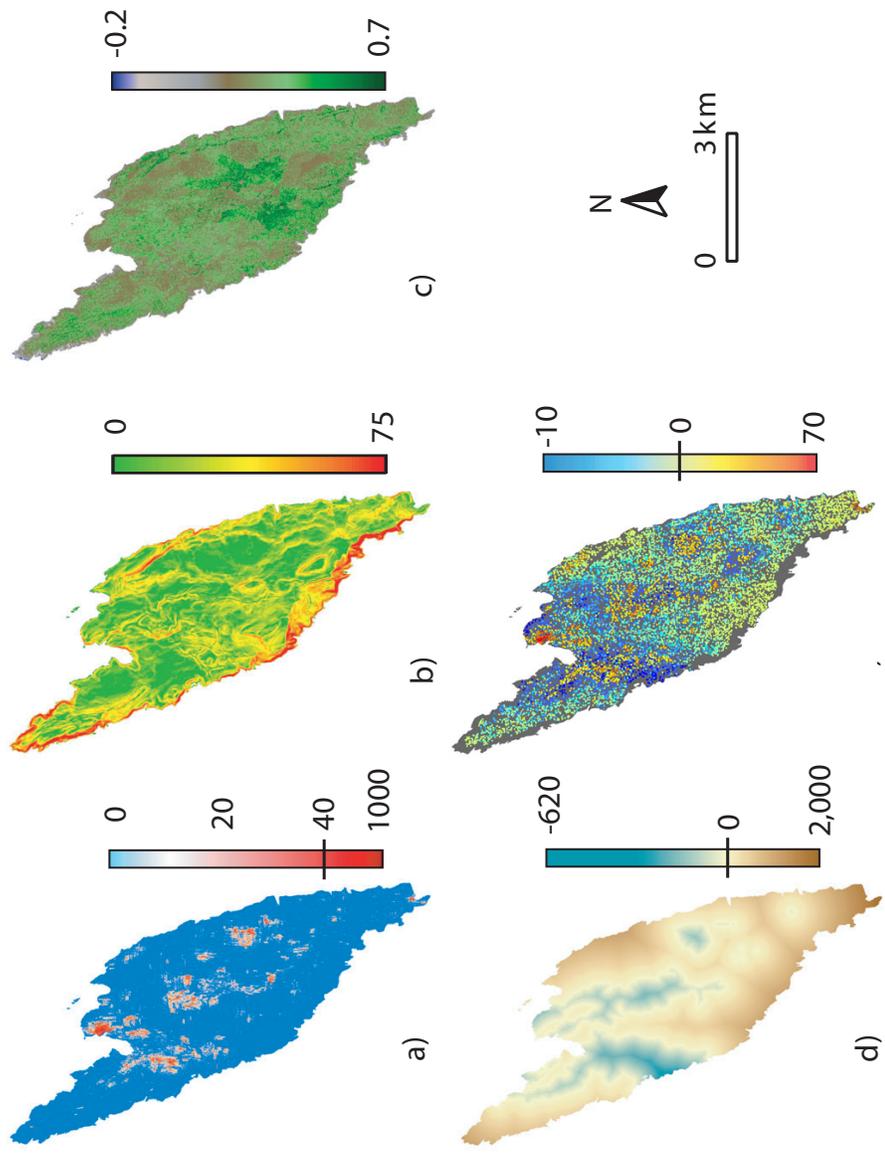


Figure 4: Variables used in a global regression of surface pottery from Antikythera: (a) pottery counts per 100 sq.m as the dependent variable, (b) slope in degrees, (c) a Normalised Difference Vegetation Index (NDVI), (d) distance in metres into (and away from) bedrock units of harder limestone, as well as (e) the deviance residuals.

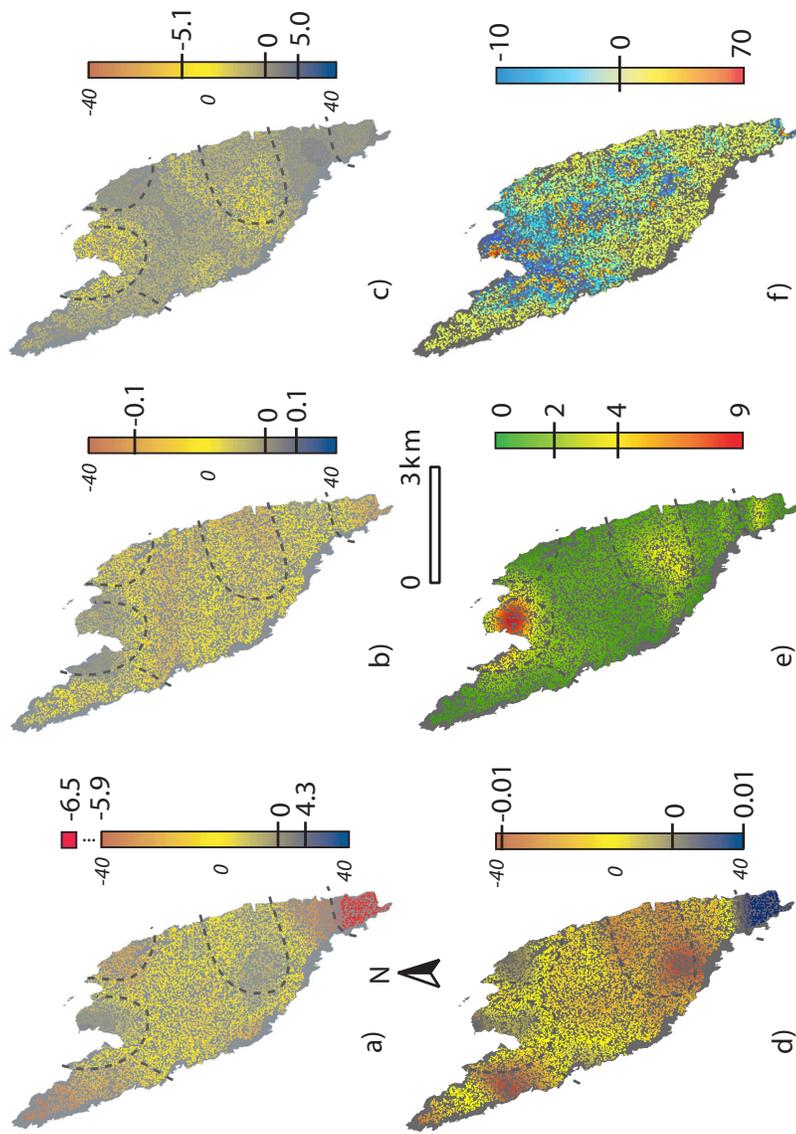


Figure 5: A geographically weighted regression of surface pottery from Antikythera showing: (a) the coefficient for slope, (c) the coefficient for NDVI, and (d) the coefficient for geology, along with (e) the local estimate of dispersion (>1 is over-dispersed and <1 under-dispersed with respect to an ordinary Poisson distribution) and (f) the deviance residuals (colour ramp is matched to fig.4e). Smaller, italicised values in the key express the number of standard errors away from the global regression estimate (see table 1). Suggested analytical sub-regions marked with dotted lines.