How Can Bayesian Smoothing and Correspondence Analysis Help Decipher the Occupational Histories of Late-eighteenth Century Slave Quarters at Monticello?

Fraser Neiman and Karen Smith—Monticello Department of Archaeology

Introduction

Two problems hinder effective intrasite spatial analyses:

1. Small samples from individual quadrats hide patterns in artifact-type frequencies in a sea of sampling variation.

2. The meaning of quadrat groups—created by clustering algorithms on the basis of similarity in type frequencies—often is opaque.

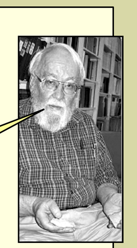
In this poster, we build on earlier work (Robertson 1999, Neiman *et al.* 2000) to explore two promising solutions: Bayesian smoothing and correspondence analysis (CA).

Bayes in Space

Bayes's theorem offers an elegant means to address the sample-size problem. Bayes's theorem shows how one can combine information about type frequencies likely to occur in a given quadrat, characterized by a "prior" probability distribution, with type frequencies actually found there to produce smoothed estimates that have lower sampling error than the raw counts.

The Bayesian estimates honor, in a statistically defensible fashion, 1) sample size in a given quadrat, 2) mean similarity of a quadrat's type frequencies to the average value for the neighborhood, and 3) mean uncertainty about type frequencies within quadrats in a neighborhood. Bayesian estimates are, therefore, superior to current methods that rely on simple weighted moving averages (*e.g.*, Neiman 1990, Whallon 1984).

The Math



A.

Consider the r quadrats that fall within the spatial neighborhood of a given quadrat. Each quadrat contains c artifact-type counts, sampled from a multinomial distribution, with unknown probabilities $\boldsymbol{\pi}_i = \{\pi_{ij}\}$ and total number of artifacts n_i. We will refer to the vector of sample proportions in the i'th quadrat as $\mathbf{p}_i$.

B.

We suppose that the unknown probabilities, from which all the quadrats in a given neighborhood are sampled, are in turn sampled from a single, "prior" Dirichlet distribution with unknown parameters β_j, which we reexpress as $K = \sum \beta_j$ and a vector of means, $\boldsymbol{\gamma} = \{\gamma_j = \beta_j / K\}$.

C.

Given the Dirichlet prior, with parameters K and γ, and a particular set of data, p_i, Bayesian estimates of the quadrat probabilities can take the form:

$$\hat{\boldsymbol{\pi}}_i = \left[\frac{n_i}{n_i + K}\right]\mathbf{p}_i + \left[\frac{K}{n_i + K}\right]\boldsymbol{\gamma}$$

Fienberg and Holland (1972, Bishop *et al.* 1975) showed that estimates like this have minimum mean-squared error when

$$K = \frac{(1 - \sum \pi_j^2)}{\sum (\gamma_j - \pi_j)^2}$$

D.

Adapting their arguments to the spatial case, we estimate the γ_j for a spatial neighborhood as the means of the quadrat proportions:

$$\hat{\gamma}_j = \frac{\sum_{i=1}^{r} p_{ij}}{r}$$

To estimate the parameter K for a neighborhood, we use the mean of r estimates of K, based on the sample proportions in each quadrat:

$$\hat{K}_i = \frac{(1 - \sum p_{ij}^2)}{\sum (\hat{\gamma}_j - p_{ij})^2}$$

Correspondence Analysis (CA)

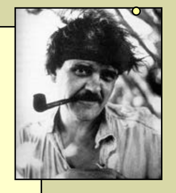
Spatial variation in artifact-type frequencies likely is caused by both temporal *and* social variation. Common practice in archaeological spatial analysis, based on cluster analysis, confounds these dimensions of variation. CA offers a means to disentangle them.

CA and Frequency Seriation

The frequency-seriation model stipulates that artifact-type frequencies arrayed in time display battleship-shaped, or Gaussian, response curves, provided the requirements of the seriation model are met.

CA and frequency seriation are intimately related. If type frequencies follow Gaussian response curves with homogeneous variances and assemblages are uniformly distributed in time, the scores of assemblages on the first CA axis approximate maximum-likelihood estimates of their temporal positions. If type frequencies have Gaussian responses to a second, synchronic gradient (orthogonal to time), the assemblage scores on the second CA axis approximate maximum-likelihood estimates of their positions on the second gradient. Hence, CA is precisely the analytic tool we need to dissect temporal and social gradients underlying spatial variation in type frequencies.

What the &%$## is a local area, anyway?



For a given seriation to monitor the passage of time, assemblages must be:

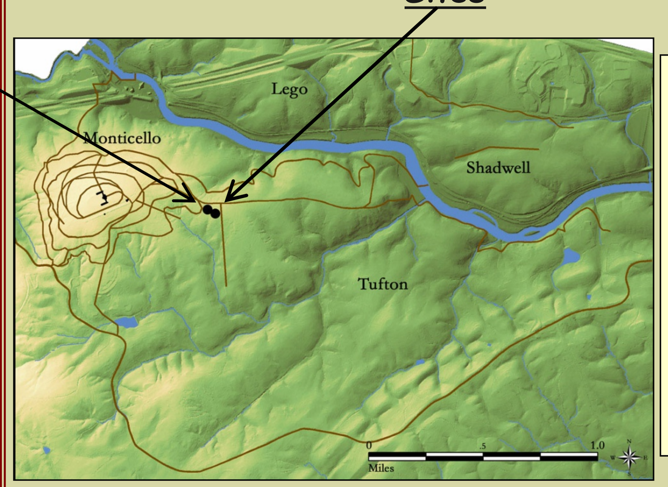
- of similar duration.
- from the same cultural tradition.
- from the same local area.

Site Background

Our case study revolves around two adjacent sites on Monticello Mountain, occupied by slaves and an overseer during the second half of the 18th century. We tested the plowzone using a stratified-random sample of 5-foot quadrats, followed by more intensive plowzone sampling adjacent to quadrats with high artifact densities or features. For more see Bon-Harper and Wheeler (2005).

Site 7

Site8



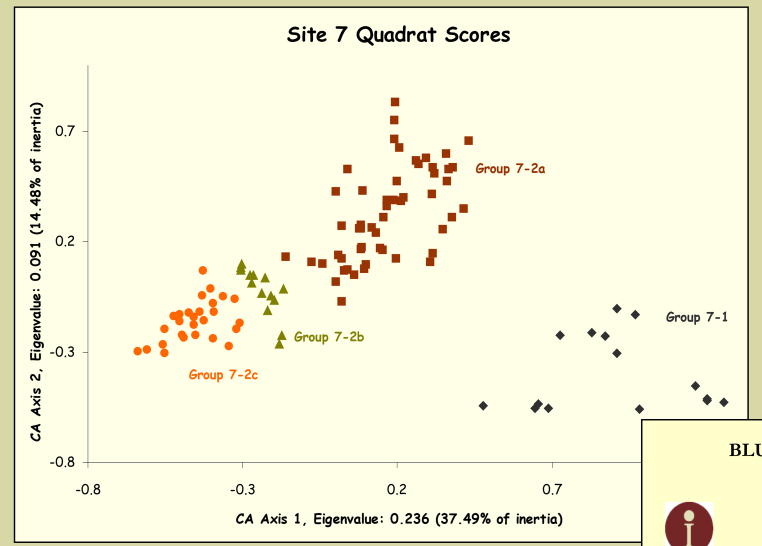
Site 7 and 8 were part the Monticello Plantation home farm. Thomas Jefferson began development of Monticello Plantation about 1770. Jefferson's father, Peter, had established a small outlying quarter farm on the mountain 30 years earlier.

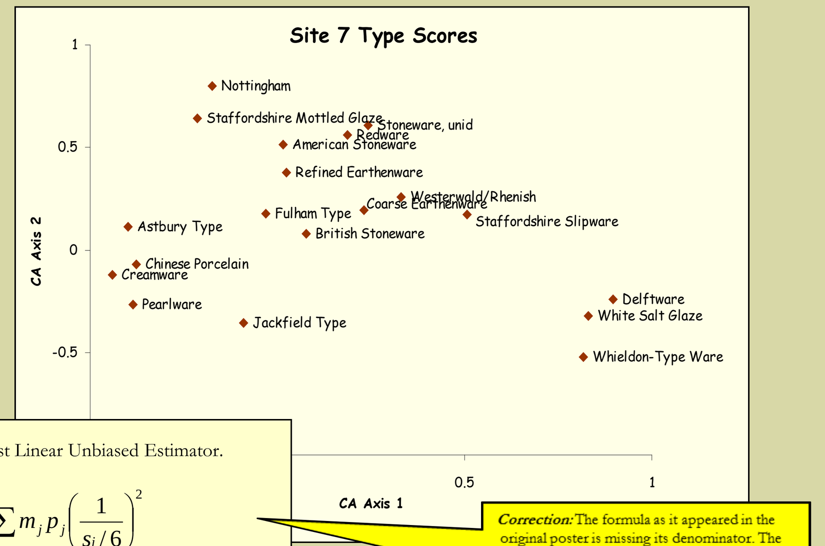
Site 7 Analysis

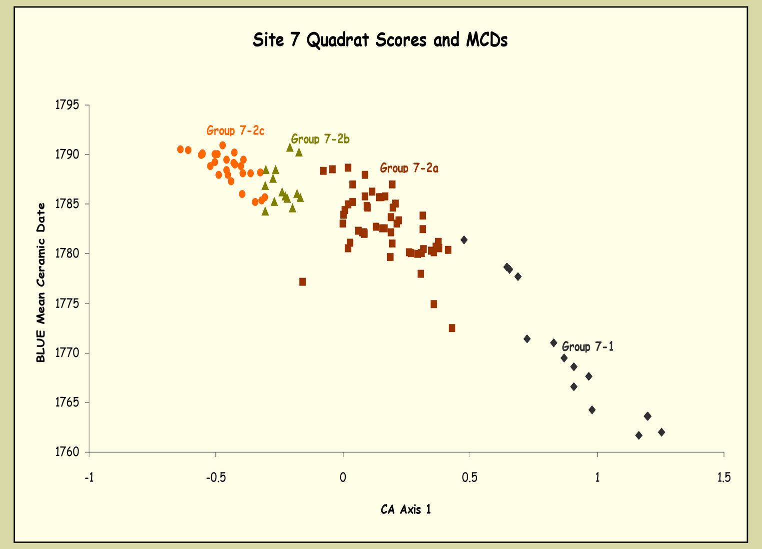
We computed Bayesian estimates of type frequencies in each 5-foot quadrat using neighborhoods with a 40-foot radius. CA suggests there are two major groups of assemblages (7-1, 7-2), the second of which was further divided into three subgroups (7-2a, 7-2b, 7-2c).

The type scores indicate that Axis 1 captures time, with early types on the right and late types on the left. Axis 2 may represent synchronic variation in cost, with cheaper ware types at the top and more expensive ones at the bottom.

We evaluated the hypothesis that Axis 1 represents time by computing BLUE mean-ceramic dates (MCDs) for each assemblage. The correlation with Axis-1 scores is strong.







BLUE is geek speak for Best Linear Unbiased Estimator.

$$MCD_{BLUE} = \sum m_j p_j \left(\frac{1}{s_j/6}\right)^2$$

where m_j is the manufacturing midpoint of the j'th type, p_j is its relative frequency, and s_j is its manufacturing span.

Correction: The formula as it appeared in the original poster is missing its denominator. The correct formula is:

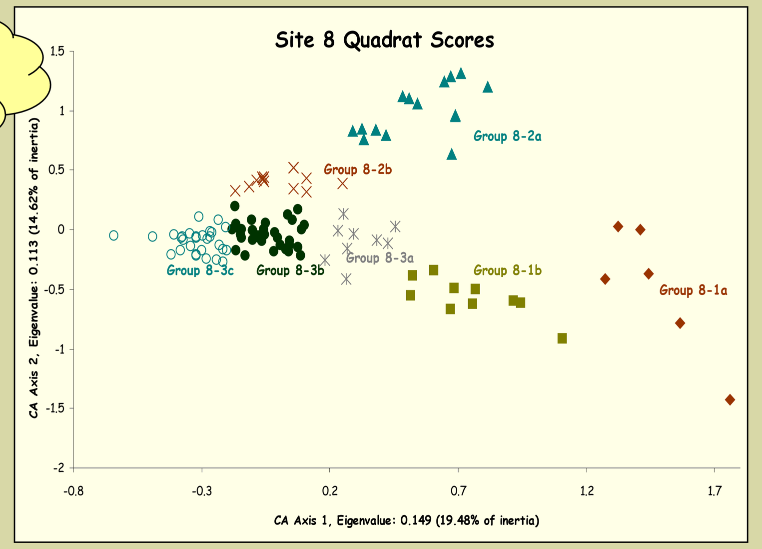
$$MCD_{BLUE} = \frac{\sum_{j=1}^{t} m_j p_j \left(\frac{1}{s_j/6}\right)^2}{\sum_{j=1}^{t} p_j \left(\frac{1}{s_j/6}\right)^2}$$

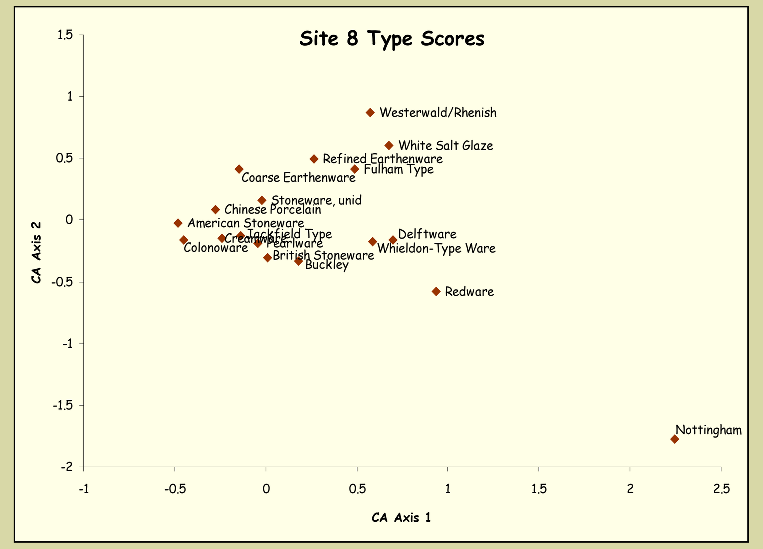
Site 8 Analysis

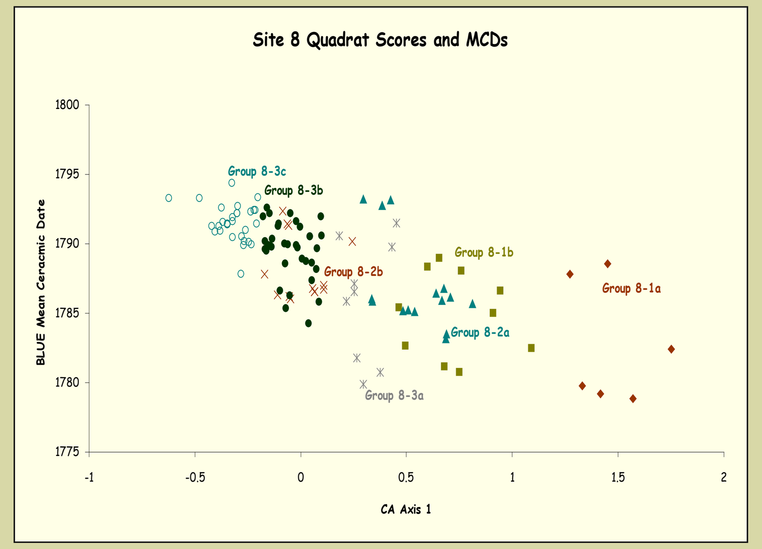
The CA of Site 8 assemblages produced a point scatter in the shape of a sideways *Y*. We assigned the assemblages to three major groups, one in each arm of the *Y* (8-1, 8-2, 8-3), and then split each group in two (a, b).

The type scores again indicate time runs from left to right along Axis 1. However, here there are unlikely to be cost differences among the types associated with Axis 2.

As at Site 7, the correlation between the BLUE MCDs and Axis-1 scores confirms that the latter captures time.







Synthesis

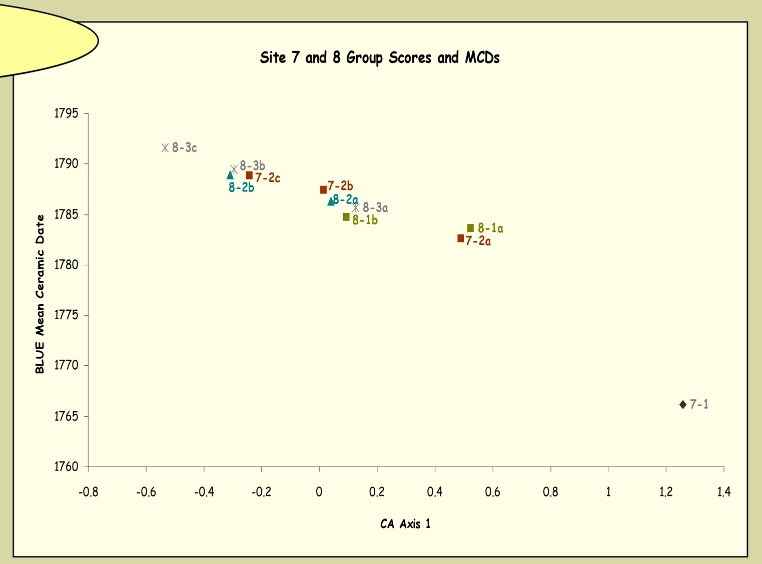
How do the assemblage groups relate to one another in time and social space? Temporal relationships among them are summarized AND confirmed by plotting Axis-1 scores against BLUE MCDs.

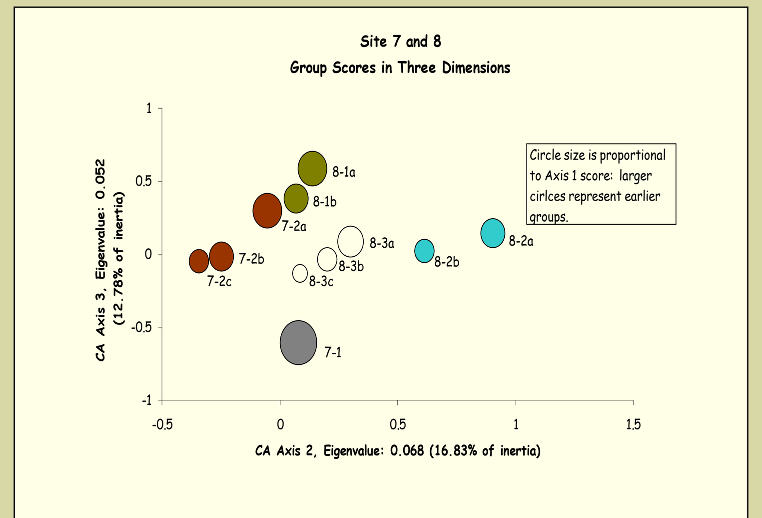
Group 7-1 is much earlier than the others. It represents the mid-18th century occupation by slaves belonging to Peter Jefferson. The remaining groups date *c.* 1770-1800 and belong to Thomas Jefferson's Monticello Plantation.

There are two additional significant dimensions of variation among the assemblage groups, captured by Axis-2 and Axis-3 scores. With the exception of 7-2a, the subgroups display historical continuity within major groups. Why is 7-2a more like 8-1a and 8-1b?

ASSEMBLAGE GROUP

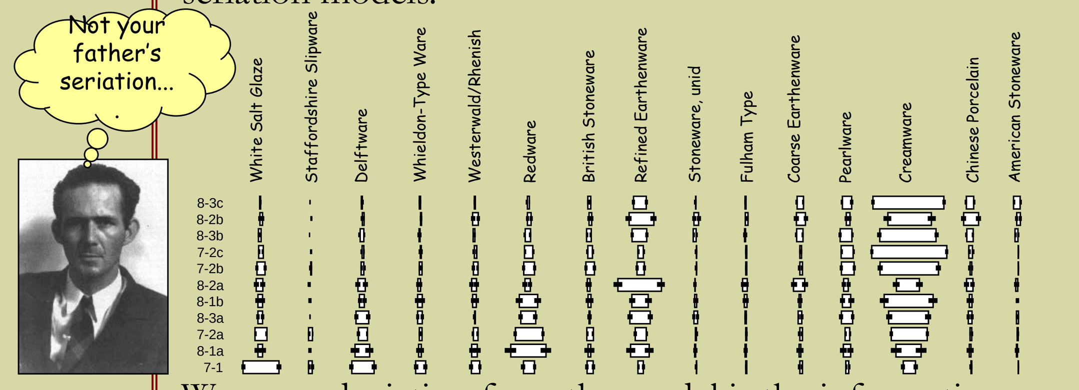
Our current guess is that an assemblage group represents a time-averaged deposit created by a group of people with access to a common suite of ceramics, whose composition is changing over time.





Discussion

If the grouped assemblages are sorted on their Axis-1 scores, the type frequencies roughly approximate the Gaussian response curves of the CA and frequency-seriation models.

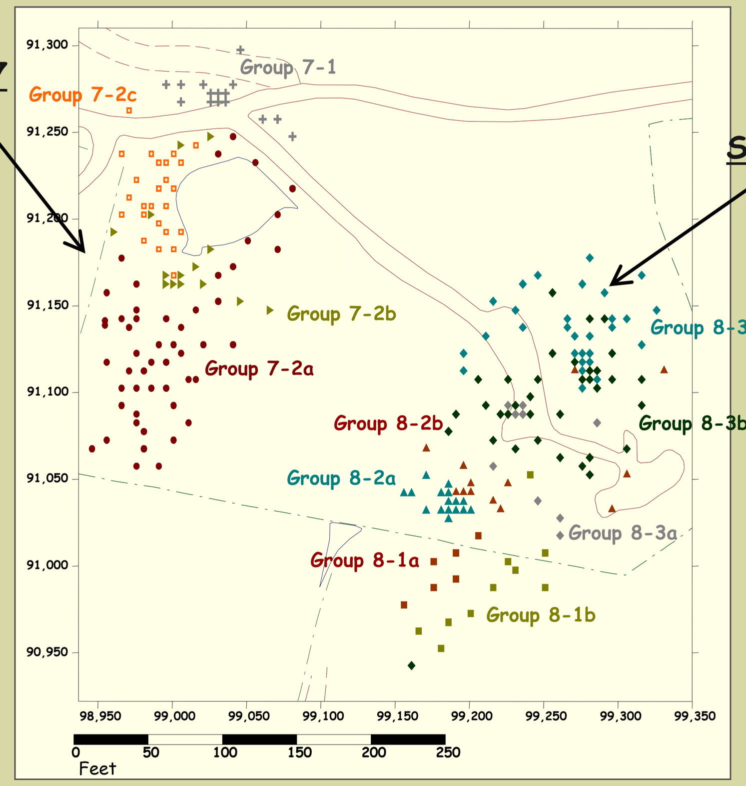


We argue deviation from the model is the informative result of different positions along synchronic dimensions of ceramic-ware abundance, especially at Site 8. Might assemblage groups represent residential groups?

In plotting the physical locations of assemblage groups, we see that 7-1 corresponds with a rock chimney base of the mid-18th century slave dwelling, whereas 7-2b and 7-2c match the location of the overseer's house (*c.* 1770-1800). We wonder if 7-2a represents a group of slaves to the south of the overseer. The affinities between 7-2a, 8-1a, and 8-1b support this idea and indicate the group moved from Site 7 to Site 8. Thereafter, two additional residential groups were established at Site 8, 8-2 and 8-3, while deposition represented by 8-1 ended. By the end of the Site 8 occupation, only 8-3 remained.

Site 7

Site 8



References

Bishop, Y., S. Fienberg, P. Holland (1975) *Discrete Multivariate Analysis: Theory and Practice*. The MIT Press, Cambridge.

Bon-Harper, S. and D. Wheeler (2005) Site Characterization: The Definition of Archaeological Sites using Plowzone Excavation Data. Poster presented at the Society for American Archaeology.

Fienberg, S. and P. Holland (1972) On the choice of flattening constants for estimating multinomial probabilities. *Journal of Multivariate Analysis* 2:127-134.

Neiman, F. (1990) An evolutionary approach to archaeological inference: Aspect of architectural variation in the 17th-century Chesapeake. Ph.D. dissertation, Yale University.

Neiman, F., S. Bon-Harper, L. McFaden, and D. Wheeler (2000) Dissecting Plowzone Palimpsests with Bayesian Spatial Smoothing and Correspondence Analysis. Poster presented at the Society for American Archaeology.

Robertson, I. (1999) Spatial and Multivariate Analysis, Random Sampling Error, and Analytical Noise: Empirical Bayesian Methods at Teotihuacan, Mexico. *American Antiquity* 64:137–152.

Whallon, R. (1984) Unconstrained clustering for the analysis of spatial distributions in archaeology. In *Intrasite Spatial Analysis in Archaeology*, edited by Harold J. Hietala, pp. 242-277. Cambridge University Press, Cambridge.

Acknowledgements

We thank Sara Bon-Harper and Derek Wheeler for collaboration in this research, Leah Stearns for her graphical wizardry (and humor), and Carl Lipo for a key photograph.