Assessing the Accuracy of Microrefuse Sampling Techniques

Isaac I Ullah, Arizona State University

SCHOOL OF HUMAN EVOLUTION & SOCIAL CHANGE

1 Introduction and Problem Statement

1) Microrefuse is commonly defined as any human-produced debris smaller than about half a centimeter.

2) The spatial patterning of microrefuse may be the most direct archaeological indicator of the locations of ancient activity areas.

3) There has been no systematic exploration of the efficiency to accuracy tradeoff of microartifact sampling procedures.

4) What are the ideal characteristics of the optimum microrefuse sampling strategy?

- To accurately capture the spatial patterning of microartifacts from an archaeological surface. • To do so using the smallest number of sample locations and to collect the lowest possible volume of soil so as to minimize costs and reduce sorting error. • To be easy for archaeologists to implement in the field so as to streamline sample collection during an excavation. • To attempt to also provide accurate results for the spatial patterning of larger artifact size classes so as to reduce the need for redundant sampling strategies for different artifact size classes.

5) This study will test the ability of several common microrefuse sampling methodologies to meet these efficiency/accuracy criteria.

- Create "virtual" microartifact and macroartifact deposits in a GIS. • Use the geostatisitcal abilities of GIS to assess the accuracy of the different sampling strategies to capture the patterning of those virtual deposits.

2 GIS Simulation of Archaeological Deposits

1) Create two distributions of "activity areas" on a 10 by 15 meter "virtual living surface" (A-D).

- A small number of large areas (A and B). • A larger number of smaller areas (C and D).

2) "Seed" artifacts in these areas (A-D).

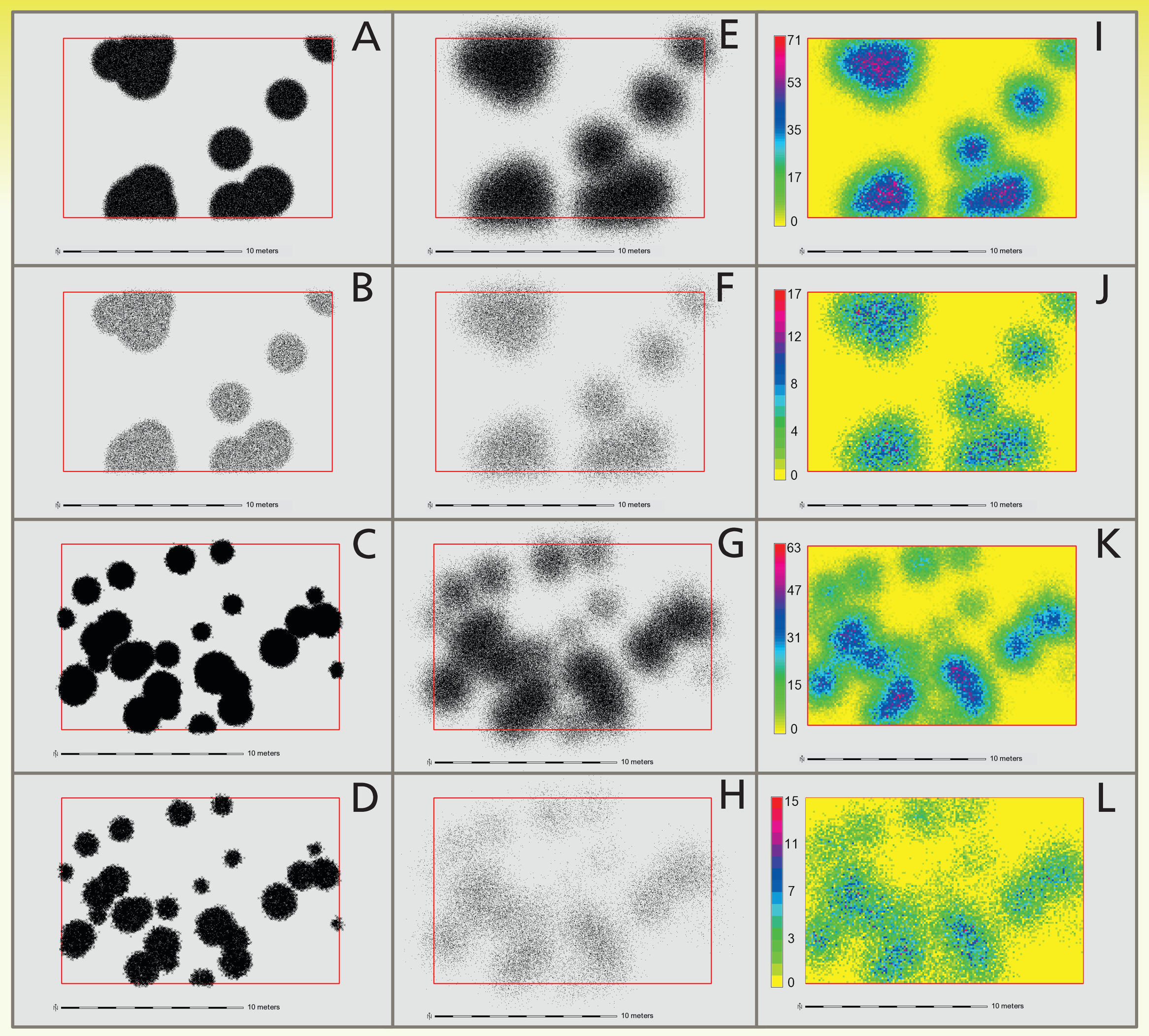
- Microartifacts at a density of 1/cm2 (A and C). • Macroartifacts at a density of 0.1/cm2 (B and D). • Artifacts distributed randomly following a normal distribution..• Activity areas are denser at their centers than at their edges.

3) Simulate site formation processes by "perturbing" the eastings and northings of each artifact (E-H).

- Distance of artifact movement determined by a random number generator following normal distribution with mean of zero.

4) Create raster density maps to act as "archaeologically real" control models (I-L).

- These maps are rasterized versions of the actual spatial pattern of artifacts on the "virtual living surface". • The maps are created at 10cm2 cell resolution. • All interpolated maps (see section 4) will be compared to these control maps.



3 Experimental Sampling Protocols

1) Standard excavation method (A, grid squares).

- One meter square grid cells. • Count artifacts in each grid cell. • Convert 1m resolution raw counts to 10cm resolution density values. • But keep a 1m spatial resolution.

2) Standard excavation method plus interpolation (A, center points of grid squares).

- Assign the 10cm adjusted density values from the 1m square grid cells (above) to a point in the center of each cell. • Use these values for interpolation (see section 4).

3) Square grid of points (B).

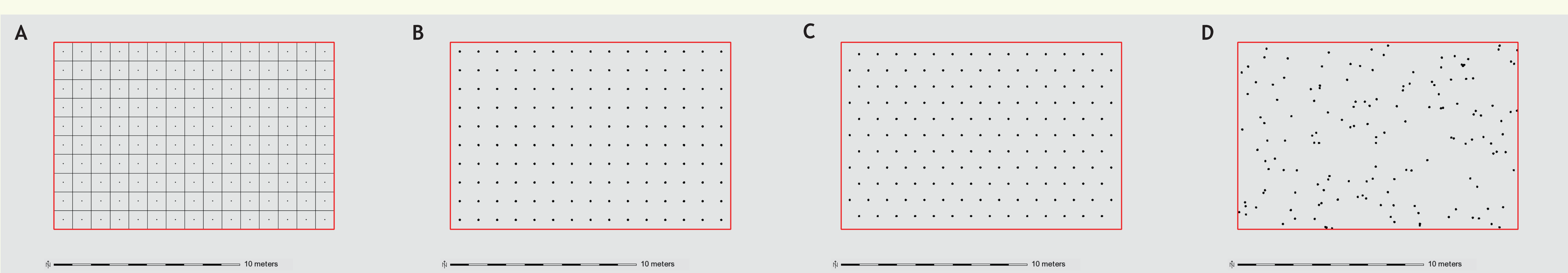
- Create a grid of sampling points where each point is a vertex in a one meter square grid. • Count all artifacts in a 5cm radius neighborhood around each point.

4) Triangular grid of points (C).

- Create a grid of sampling points where each point is a vertex in a one meter triangular grid. • Count all artifacts in a 5cm radius neighborhood around each point.

5) Random points (D).

- 150 sample points randomly located across the surface. • Count all artifacts in a 5cm radius neighborhood around each point.



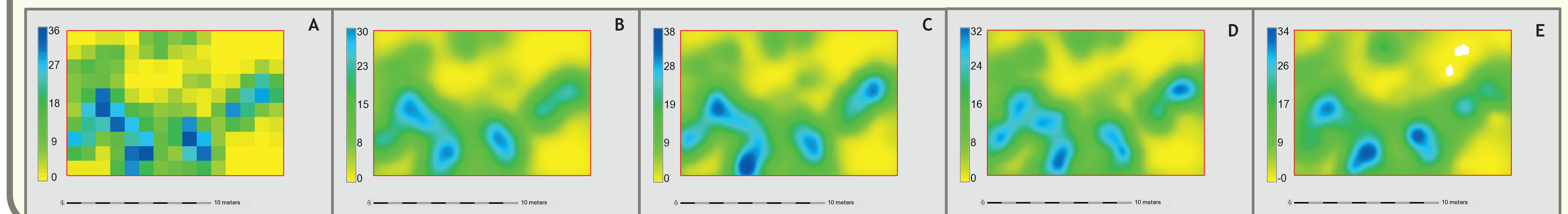
4 Interpolation of Raster Density Maps

1) Sampled density data turned into 10cm resolution raster density maps using Regularized Spline Tension Interpolation.

- Interpolation statistically "fills the gaps" between each input point. • Predicts the spatial pattern of artifact densities at a finer scale than of the original collection grid.

2) Each Sampling procedure produces a different interpolated result

- Examples here are interpolated results from sampling a dense microrefuse spread over a large number of small activity areas (see Section 2, figures C, G, and K):
- A) Raw one meter grid counts. • B) Interpolated from one meter grid counts. • C) Interpolated from a rectilinear grid of points. • D) Interpolated from a triangular grid of points. • E) Interpolated from random points. • Color scales for figures A-E match the actual density map in section 2, figure K.



5 Results

1) Accuracy of density value predictions (Table 1).

Assessed by comparing the covariance and regression coefficients of each interpolated map against the real 10cm density maps (Table 1).

- Interpolating from the grid counts produced the most accurate density predictions.
- Interpolating from a triangular or rectangular grid also produced accurate density predictions.
- Interpolating from random points and the raw grid counts produced less accurate density predictions.
- The case of sparse macrorefuse spread over many small areas provides the one exception to this pattern.

2) Accuracy of activity area shapes (Charts 1 and 2, figures A-D).

Assessed by the difference between binary presence/absence maps of the control and interpolated models (see examples A-D). The percentage of area under and over-predicted by each sampling strategy are compared (Charts 1 and 2).

- Interpolating from the one meter grid counts gave the most spatially accurate results for microrefuse (A and B).
- Interpolating from the square grid of points produced the best results for macrorefuse (C and D)

Table 1

A) Microrefuse – 10 Areas

	Covariance	R-squared
Grid Cells Raw Data	0.925	0.856
Interpolated From Grid Cells	0.956	0.914
Square Grid Points	0.943	0.840
Triangular Grid Points	0.945	0.893
Random Points	0.931	0.803

B) Microrefuse – 30 Areas

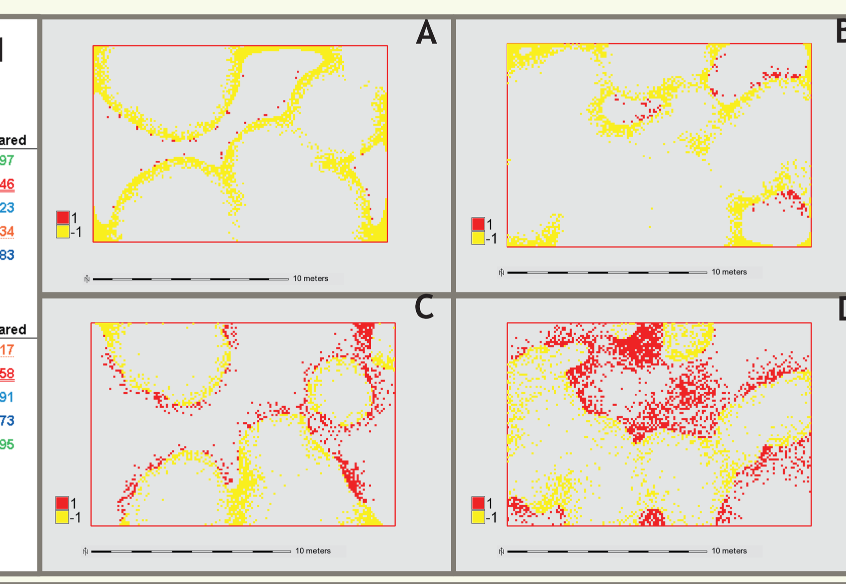
	Covariance	R-squared
Grid Cells Raw Data	0.893	0.797
Interpolated From Grid Cells	0.920	0.846
Square Grid Points	0.907	0.823
Triangular Grid Points	0.913	0.834
Random Points	0.885	0.783

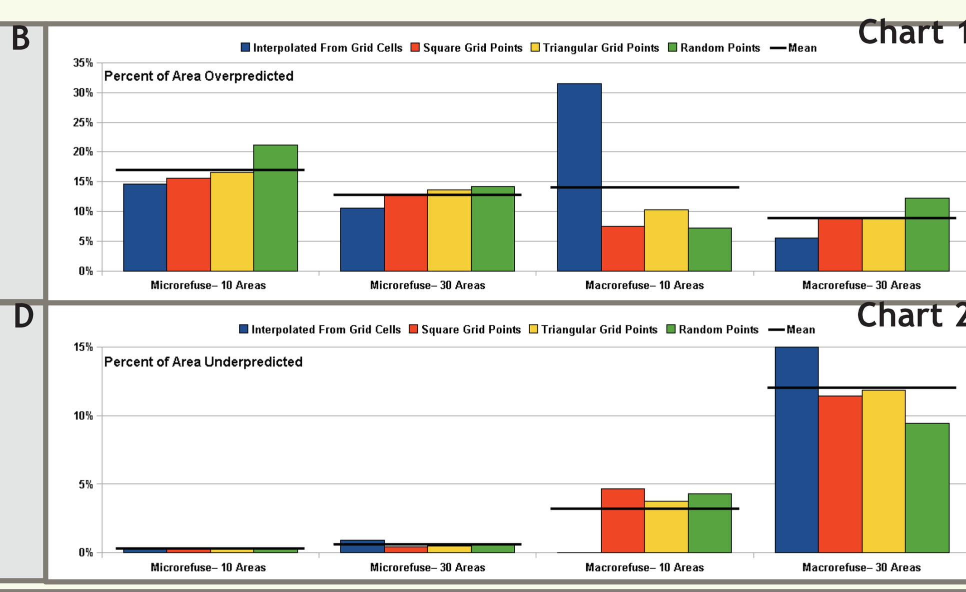
C) Macrorefuse – 10 Areas

	Covariance	R-squared
Grid Cells Raw Data	0.822	0.675
Interpolated From Grid Cells	0.853	0.727
Square Grid Points	0.826	0.683
Triangular Grid Points	0.829	0.687
Random Points	0.820	0.673

D) Macrorefuse – 30 Areas

	Covariance	R-squared
Grid Cells Raw Data	0.719	0.517
Interpolated From Grid Cells	0.747	0.558
Square Grid Points	0.701	0.491
Triangular Grid Points	0.688	0.473
Random Points	0.704	0.495





6 Conclusions and Future Directions

1) There are significant differences between the tested microrefuse sampling strategies

2) But no single strategy worked the best for all activity area configurations for both artifact size classes.

3) Interpolating from density adjusted one meter grid cell counts produces the most accurate density values.

- This is probably because the neighborhood size is large enough to even out random "spikes" in the actual density. • However, this method would require the most amount of soil to be collected, and is thus is the most costly and time consuming method. • It also does not always accurately predict the shape of activity areas.

4) Both the triangular and square grid of sampling points gave accurate density results.

- These methods also best preserved the shape of the activity areas. • They are faster to set up in the field, and require a much smaller amount of soil to be collected, and are thus the most cost effective. • However, these methods do not work as well for sparse macrorefuse deposits.

5) Thus, the "optimum" solution seems to be to superimpose a square grid of microartifact sampling points on a cellular sampling macroartifact sampling grid.

- After microartifact soil samples are removed at the grid points for later laboratory analysis, the rest of the grid cell can be excavated and screened in the field through a large mesh screen.

6)Future experiments:

- Test a larger variety of activity area configurations (potentially including ethnoarchaeologically and/or experimentally derived activity area distributions). • Test the effect of larger collection radii around sampling points • Test different sampling grid spacing/resolution. • Implement a Monte Carlo simulation of artifact points (multiple re-randomization) to avoid stochastic errors.