Multi-Scale and Nested-Intensity Sampling Techniques for Archaeological Survey

Oskar Burger
The University of New Mexico
Albuquerque, New Mexico

Lawrence C. Todd
Paul Burnett
Colorado State University
Fort Collins, Colorado

Tomas J. Stohlgren
U. S. Geological Survey
Fort Collins Science Center, Fort Collins, Colorado

Doug Stephens
U. S. Forest Service
Chadron, Nebraska

This paper discusses sampling techniques for archaeological survey that are directed toward evaluating the properties of surface artifact distributions. The sampling techniques we experimented with consist of a multi-scale sampling plot developed in plant ecology and the use of a nested-intensity survey design. We present results from the initial application of these methods. The sampling technique we borrowed from plant ecology is the Modified-Whittaker multiscale sampling plot, which gathers observations at the spatial scales of 1 sq m, 10 sq m, 100 sq m, and 1000 sq m. Nested-intensity surveys gather observations on the same sample units at multiple resolutions. We compare the results of a closely-spaced walking survey, a crawling survey, and a test excavation to a depth of 10 cm. These techniques were applied to ten 20 x 50 m survey plots distributed over an area of 418 ha near the Hudson-Meng Bison Bonebed in nw Nebraska. These approaches can significantly improve the accuracy of survey data. Our results show that high-resolution coverage techniques overlook more material than archaeologists have suspected. The combined approaches of multi-scale and nested-intensity sampling provide new tools to improve our ability to investigate the properties of surface records.

Introduction

Archaeological survey has changed dramatically over the years. While at one time the need to survey had to be justified, the concern with regional patterning has continually developed and today survey is among the most fundamental techniques of archaeological inquiry (Ammerman 1981; Banning 2002; Schiffer, Sullivan, and Klinger 1978). The goals of survey projects and the demands placed on survey data have become more diverse and complicated with time. The observations made by crews walking parallel lines over the landscape must address issues ranging from finding and protecting sites from damage to investigating the nature of past land-use strategies. Because of the diversity of potential applications of survey data and the challenges associated with obtaining and understanding them, multiple concerns may compete for priority in the selection of field techniques. Surveys may need to cover very large areas in little time as well as account for the effects of a variety of taphonomic factors that influence the regional record. Ideally, the data supporting interpretations and management decisions are efficiently gathered but also of sufficient quality to accurately represent the distributional properties of the regional record.

Both management and research perspectives (although these need not be mutually exclusive) can benefit from large spatial samples that reveal associations between ge-
ography and material culture and a thorough understanding of the processes that influence the distributions that surveys document. Accordingly, the numerous goals of archaeological survey are here grouped into discovery-based and property-based modes of investigation. Discovery-based surveys identify geographical aspects of the surface record by locating and describing clusters of artifacts. A property-based approach focuses on evaluating accuracy of method and technique and formational aspects of the regional record (Ebert and Kohler 1988; Shott 1995; Wandsnider and Camilli 1992). The capacity to analyze and interpret the contexts of discovery could be improved by conducting more detailed analyses of the factors that influence the surface samples that we use to infer spatial relationships. This requires embracing an explicitly experimental approach and integrating property-based investigation as one of several phases in a survey project (Given et al. 1999; Schiffer, Sullivan, and Klinger 1978). Siteless or distributional surveys have made major contributions in this regard (Ebert 1992; Foley 1981; Thomas 1975). By sampling regions as opposed to “sites,” these surveys attempt more holistic and accurate interpretations of landscape records.

We present two additions to the conventional survey toolkit that are aimed at developing property-based investigation. The first is the use of a multi-scale sampling plot developed in plant ecology. The second is a nested-intensity survey design that covers each plot with more-than-one observer resolution. These techniques were applied in an archaeological survey on the Oglala National Grassland (ONG) of NW Nebraska (Fig. 1). A property-based approach is valuable for investigating geomorphologically active and topographically diverse archaeological landscapes like the ONG (Fig. 2), but the basic principles are applicable to many other sampling situations whether landscape-oriented or intra-site.

Figure 1. Location of the Oglala National Grassland, Nebraska within the United States.
The technique we borrowed from plant ecology is the Modified-Whittaker multi-scale vegetation sampling plot (FIG. 3). This sampling design was applied to archaeological survey because of the increased accuracy of the plant diversity samples gathered with this method, the experimental control that the plot provides, and because of its multi-scale design. Plants and artifacts share some basic distributional properties, so it follows that a survey design that is highly useful in plant species surveys may also be useful for sampling artifact distributions. "Artifacts share many properties with plants, in having small unit size in relation to a very large spatial context, and also by having a patchy distribution" (Foley 198l: 174). The size of the Modified-Whittaker plot is 1000 sq m, which is a very small area compared to conventional designs. Orton (2000: 88) considers survey units that are 500 to 1000 m on a side as "small" archaeological sample units. Thus, we were initially hesitant to apply the Modified-Whittaker plot directly to archaeological survey. As we became more aware of the improvement in the accuracy and reliability of plant species surveys conducted with this method, however, we decided to test its utility in an archaeological case study.

An additional methodological issue came to light when one of our colleagues in plant ecology wanted to know how archaeologists estimate the number of artifacts they miss, or walk past, when conducting a pedestrian survey. We realized that we had no way to reliably answer this question with conventional techniques (Banning 2002: 62) and thus we elected to conduct a nested-intensity survey to address this issue. A nested-intensity survey consists of covering the same sample units with different resolutions. We covered each of the 10 Modified-Whittaker plots in this study with a fine-grained walking survey, a crawl survey, and then conducted test excavations of the top 10 cm. Comparing the results of the samples obtained by each technique provides a means for investigating the effects of basic decisions on the estimates concerning the overall population. The intensive coverage provides high-resolution samples of limited areas that can be used to evaluate the accuracy of our approach.

Before discussing the survey in greater detail, we should emphasize that the techniques of multi-scale and nested sampling are strategies for systematic sub-sampling that can be used to augment traditional survey practice. These
are not considered replacements for conventional techniques that gather coarser-grained samples of larger spatial extent. Certain aspects of survey certainly require large coarse-grained samples. The small area covered by this approach could not be used on its own to achieve an understanding of a regional record but this small-scale sampling plot is ideal as an experimental framework for property-based investigations. The nature and significance of the items found by discovery-based surveys could be placed into more informative contexts if they were complemented with high-resolution samples aimed at evaluating the properties of the record itself. The techniques we present are intended to enhance discovery by focusing on the evaluative goals of survey which are often not given as much emphasis in the design of regional sampling schemes.

The Modified-Whittaker Plot and Multi-Scale Sampling

The Modified-Whittaker multi-scale sampling plot is an improvement over conventional rangeland techniques for studying plant diversity (Stohlgren, Falkner, and Schell 1995; Stohlgren, Bull, and Otsuki 1998). In comparative studies, the Modified-Whittaker plot outperformed traditional plant sampling designs by documenting more plant species, capturing more rare species, and more accurately representing the relative abundances of species in the community. The traditional techniques that typify plant species-richness surveys are analogous to those in archaeology. Conventional plant survey methods involve documenting species within quadrats placed along a transect, whereas archaeological surveys consist of long linear transects walked.
by surveyors who continually search the ground surface for cultural material. Unlike plant survey techniques, archaeological surveys are often conducted without a specific regard for spatially defined sample units. The transect-based plant sampling methods over-represented the dominant plants while failing to capture representative proportions of non-native or rare species (the many species that each contribute less than 1% of the total vegetation cover) and often missed them altogether (Stohlgren, Bull, and Otsubo 1998: 168). The transect methods were also poor predictors of the total number of species, missing about half of the native and non-native plant species in all habitat types. Generally, similar observations can be made of archaeological survey methods.

Wandsnider and Camilli (1992) demonstrated that a similar bias often results from pedestrian transect methods used in archaeological survey, which tend to over-represent dominant parts of the surface record. Wandsnider and Camilli's (1992: 184) investigation of survey accuracy found that "survey with an acceptable transect interval of 15 m will intercept, at most, 6-13% of the members of a low-density artifact population, but only some of these items will actually be found." A transect spacing of 5 m was more likely to find clustered rather than isolated artifacts. In an experiment involving seeded artifacts (painted washers and nails), crews recovered 82% of the clustered material compared to just 16% of the isolated material (Wandsnider and Camilli 1992). It is therefore likely that traditional pedestrian surveys result in shaky inferences about landscape patterns due to a lack of understanding of the properties of the sample.

The improved performance of the Modified-Whittaker plot over conventional plant sampling is based on the spatial layout of the subplots (FIG. 3a). The 0.5 x 2 m subplots (FIG. 3a, numbered 1 to 10) are arranged to reduce the amount of spatial autocorrelation in the sample, which is the principle that two points close to one another are more likely to be similar than two points that are further apart. Within the 1000 sq m plot (K), these 1 sq m subplots are arranged to improve the accuracy of their spatial coverage. In archaeology, this is useful as a method for gathering systematic small-scale samples of the surface record and also offers an ideal arrangement for test excavations. Standard pedestrian transects, which are essentially "numerous quadrats placed end-to-end" (Orton 2000: 90), have a high degree of spatial autocorrelation. The Modified-Whittaker plot's multi-scale design can be used to counter this problem and results in more representative samples (Stohlgren, Falkner, and Schell 1995).

The incremental increase in the spatial scales of the subplots within the Modified-Whittaker frame leads to its increased utility for evaluating population variables. The ten 1 sq m subplots are placed along the internal boundary of the 1000 sq m plot and along the external edge of the 100 sq m C subplot (FIG. 3a). Subplots A and B are 2 x 5 m in dimension (10 sq m) and are in opposite corners of the Modified-Whittaker plot. The central rectangle, plot C, covers an area of 100 sq m and measures 5 x 20 m (FIG. 3a). The exact distances between subplots are presented in Figure 3b, which is based on tracing the perimeter of the plot with a 100 m tape measure anchored at the lower right corner of the plot. All of the subplots within the K plot are non-overlapping so that the properties of each sampling unit can be evaluated independently (Stohlgren, Falkner, and Schell 1995). This allows the properties of the sampled population to be analyzed with respect to changes in spatial scale. Combining this design with siteless survey may aid in eliminating the bias of favoring obtrusive and dense portions of the archaeological record.

Accurate Samples as Tools for Evaluating Discovery

One major value of a nested-intensity survey is that it provides archaeologists with the means to evaluate the accuracy of a surface sample in terms of the number of items that surveyors did not discover. High-resolution coverage can be used to investigate how the variables of the sample deviate from those of the population. "[T]he systematic acquisition of the whole archaeological record, including the low-density areas, can yield more valid and reliable insights into the nature of the archaeological record, even the nature of archaeological sites" (Dunnell and Dancey 1983: 274). Likewise, Cowgill pointed out that the "most obvious way to get relatively good information on even very tiny occurrences (insofar as they are visible at all on the surface) is to increase the intensity of survey by spacing survey team members so closely that nearly all tiny occurrences will be spotted" (1990: 257). While Cowgill was not intentionally referring to crawl surveys (personal communication, 2001), archaeologists should have an understanding for just how closely surveyors need to be spaced in order to actually find "nearly all tiny occurrences."

In order to evaluate the properties of a large spatial sample, archaeologists may find it useful to include intensively sampled survey plots, even if it means reducing their overall spatial coverage. In excavations, archaeologists have tended to sacrifice area in favor of more thorough recording strategies, producing smaller samples at higher resolution. Such fine-grained samples favor documenting context over object discovery and are based on the realization that there is more information contained in the relationships between objects and their surroundings than in the
objects themselves (Clarke 1977). Similarly, the mesh sizes used to screen archaeological deposits have become progressively smaller: 1/4-inch dry screen was once the norm and now many archaeologists water screen at least a portion of their deposits through 1/16-inch mesh. Transect spacing can be considered as analogous to screen sizes in the context of survey methods. Traditional techniques have focused on the highly dense aspects of the record, while much of the rest has slipped by in the gaps between transects.

The ONG Survey Project and Method

Our study area is the grassland surrounding the Hudson-Meng Bison Bonebed in NW Nebraska. The survey was conducted as part of the Colorado State University archaeological field school in cooperation with the United States Forest Service during the summers of 1998–2001 on the Oglala National Grassland (ONG Survey Project). The region consists of rolling northern mixed prairie, pine-tree crowned ridges, and badland exposures (Fig. 2). Chipped stone is by far the dominant artifact class and the prehistoric record was generally dominated by hunter-gatherer behavior patterns. The initial goals of the survey were to evaluate the utility of the Modified-Whittaker plot for archaeological survey and to investigate the effects of observer intensity on artifact recovery.

Nested-Intensity Sampling

Nested-intensity sampling covers the same section of ground at progressively finer resolutions. For the ONG surveys, the first coverage involved walking one set of parallel transects over the entire 1000 sq m K plot with about 70 cm between surveyors. The guideline for crew spacing was that each surveyor could touch the shoulder of the next person in the survey line. The second coverage consisted of a different survey team crawling, with shoulders touching, over subplots I to 10, A, B, and C (Fig. 4). The total area crawled was 130 sq m (Fig. 3A: subplots I to 10 are 1 sq m, A and B are 10 sq m, and C is 100 sq m). All discovered artifacts from both surveys were left in place and pains were made to return them to their original locations after recording. As a result, the crawling survey would rediscover many of the artifacts recorded by the walking survey. After the walking and crawling surveys were completed and all discovered artifacts had been recorded, the top 10 cm of sediment of each 0.5 x 2 m subplot (subplots 1 to 10) were excavated and passed through 1/8-inch screen.

The excavation is considered to be a measure of the population of artifacts that could potentially be exposed and discovered. The subsurface test is necessary because archaeologists often use surface artifact densities as predictors of subsurface densities, but several processes can ex-
Table 1. Summary data for documented chipped stone in all Modified-Whittaker plots. Plot locations are shown in Figure 5. Means are calculated by taking the total for each column and dividing it by the total number of subplots. For subplots 1–10, each total is divided by 100; for subplots A and B, totals are divided by 20; for subplot C and plot K, the totals are divided by 10.

<table>
<thead>
<tr>
<th>Modified-Whittaker plot</th>
<th>Subplots 1 to 10</th>
<th>Subplots A and B</th>
<th>Subplot C</th>
<th>K plot</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Walk</td>
<td>Crawl</td>
<td>Excavate</td>
<td>Walk</td>
</tr>
<tr>
<td>NRTP</td>
<td>3</td>
<td>46</td>
<td>2342</td>
<td>8</td>
</tr>
<tr>
<td>WRIN</td>
<td>0</td>
<td>1</td>
<td>66</td>
<td>0</td>
</tr>
<tr>
<td>WROUT</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>CC01</td>
<td>0</td>
<td>0</td>
<td>37</td>
<td>0</td>
</tr>
<tr>
<td>WH01</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>S3N6</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>N68S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WTRI</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>WTRO</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>TTO1</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>Mean</td>
<td>0.03</td>
<td>0.5</td>
<td>25</td>
<td>0.5</td>
</tr>
<tr>
<td>1 Standard error</td>
<td>0.0</td>
<td>0.2</td>
<td>9.0</td>
<td>0.3</td>
</tr>
</tbody>
</table>

pose artifacts near the surface, so passing the earth through a screen measures artifact density in the near-surface taphonomically active zone or TAZ (Lyman 1994: 405).

Following Wandsnider and Camilli (1992), surface items were separated according to the manner in which they were discovered. The locations of all items found during the systematic walking survey were marked with a red pin flag. Each red-flagged item was then recorded by a team of two to three people seated or kneeling around the pin-flagged artifacts. The locations of artifacts found nonsystematically, or during the recording process, were marked with a blue pin flag and also recorded. Due in part to our non-collection recording strategy, a number of attributes were documented for each item including length, width, thickness, inclination, orientation, and a variety of other descriptive codes consistent with the recording strategy used in the excavation of the Hudson-Meng Bison Bonebed (Todd 1987). As a result, the ground surface around the red-flagged (systematically discovered) artifacts was intensively scrutinized during the recording process, providing opportunities to make blue-flag (nonsystematic) discoveries. We presumed that the intensive double-coverage of the area provided by the systematic and nonsystematic coverage would lead to a nearly complete document of the surface record and implemented the crawl survey to test that assumption. We did not think that the crawl survey would add a significant number of artifacts. As discussed below, we were very wrong.

Setting Up the Modified-Whittaker Plots

Because of the elaborate internal structure of the Modified-Whittaker plot, the set-up process needs to be as efficient as possible. Over the course of our study we experimented with two approaches to plot set-up. The first was highly controlled and time-consuming in that the exact provenience of each corner of each subplot was fit to the UTM grid with precision mapping equipment, locus GPS receivers or a total station; both provide sub-centimeter accuracy. When the multi-scale plot was first used we were conducting a number of experiments with survey intensity and spatial variability that demanded such accuracy. As the project developed, however, expediency of set-up became more of a concern and a more "low-tech" and time efficient method was adopted that involves tape measures and a compass. To ensure consistency, the plots were set up according to the arrangement in Figure 3b, which is included as a guide for any future applications of this method. A Modified-Whittaker plot can be set up in under 30 minutes using the low-tech approach.

Results

Ten Modified-Whittaker plots were placed within a 418 ha area of grassland (fig. 5). No statistically-derived sampling scheme was used to determine the placement of the plots on the landscape, but special effort was made to select locations that differed in topography, ground surface visibility, surface chipped stone density, and degree of disturbance from grazing cattle. Some of the plots were placed where artifacts were known to be abundant, but others were in locations that would frequently escape archaeological analyses due to high vegetation cover or low artifact density. Table 1 shows the variability in chipped stone recovery and provides an initial summary of the relationship between method, spatial scale, and artifact density. Several of the plots yielded artifacts in subsurface tests when few or none were seen on the surface. None of the
plots were placed where a previous site had been recorded, but all plots were within the limits of previously surveyed areas. Here we focus on the results of the study’s property-based investigations and emphasize the plot’s applicability to archaeology as an experimental sampling design.

**Two Examples: Plots NRTP and WRIN**

For purposes of illustration we have selected two sample plots for detailed description. Plot NRTP was located on a prominent, flat-topped hill where a high density of surface artifacts was known to exist. Our initial focus was to evaluate how much material evades discovery from fine-grained methods in order to gain perspective on how many artifacts are overlooked by standard transects. Of the ten plots, NRTP had the highest density of surface materials recorded. In the 1000 sq m of the K plot, the walking survey found 366 items systematically (red-flag discoveries) and this number increased to 867 when the blue-flagged discoveries were included. Thus, even at this close spacing, more than half of the artifacts were found while recording after the systematic walking phase was completed. The crawl survey found 3.5 times more material (244 artifacts) than the walking survey (65 artifacts) in 130 sq m (subplots 1 to 10, A, B, and C).

The substantial increase in items found by the crawling survey is evident in the distributions of artifacts recovered.
by the different methods (FIG. 6).¹ Artifact clusters were identified during the crawl survey that were not apparent in the combined systematic and nonsystematic distributions from the walking survey (FIG. 6A–B). The nonsystematic coverage, while adding many additional finds, tended to highlight existing clusters rather than identify new ones. This is especially evident when the densities in subplot C are considered. The crawl survey transformed a relatively diffuse scatter into a dense cluster (FIG. 6C).

NRTP also had a dense subsurface artifact population.

¹ The arrangement of the subplots is slightly different for NRTP than for the other plots in the study. For NRTP, all 10 of the numbered 1 sq m subplots are on the perimeter of the K plot but for all others subplots 7–10 are placed on the edge of the C subplot as pictured in Figure 3A. This is an adjustment made to the Modified-Whittaker plot by Stohlgren, Bull, and Otsuki (1998) in order to further reduce the spatial autocorrelation among subplots. We shifted to the adjusted version in the summer of 2000. The plot depicted in Figure 3A could be referred to as the “revised Modified-Whittaker” but such terminology seems excessive.
The analysis of the data from WRIN demonstrates that samples of the surface record in this area can change in very short time periods. Because of the grazing study, WRIN was resurveyed after one calendar year had passed (along with other experimental plots) in order to compare short-term changes in range condition and other factors. The chipped stone was also re-recorded. The results of these two surveys were plotted, one on top of the other, in order to compare two documents of the same area gathered at equal intensities (FIG. 7). The chipped stone observed during the respective surveys produced two entirely different archaeological documents, although only one year had passed and the plot lies in an ungrazed setting. At this point, the discrepancies between the two samples can most likely be attributed to a combination of factors. The artifacts that individual surveyors are likely to see at a given time will vary with changes in sunlight intensity and their experience level. It is also likely that the visible surface record had altered its composition due to variables such as small-scale vegetation changes and sediment cover.

To document the effects of grazing on the movement and visibility of archaeological materials we seeded and provenienced individually numbered aluminum clasts on the surface of WRIN (and other plots) to track their movement over time. The clasts were flat rectangles and trapezoids or cylinders between 16 and 45 mm in length. The clasts had moved over the course of a year. The largest movement on WRIN was 9 cm, but almost all moved 3 cm or less. Of the 20 aluminum clasts on the surface of WRIN, 6 were not relocated by the second survey. This suggests that the differences in results between the two surveys were due to changes in the record itself, and cannot be explained by crew experience alone, because the aluminum clasts are large, shiny, and very conspicuous against the brown substrate. It is clear that the taphonomic agents that rearrange, obscure, and expose surface artifacts are active over short periods of time and in settings that appear to be relatively undisturbed. Understanding variability in these properties of the archaeological record is difficult, but these data demonstrate how property-based investigations can be used to evaluate the relative stability of the surface artifact population.

**Walking vs. Crawling**

Achieving a reliable measure of survey accuracy required a sample that was nearly equivalent to the actual surface record. While artifact size and limits of visual acuity require that items will always be missed, those documented on the crawl survey appear to approximate closely the observable surface record. No nonsystematic discoveries were made while recording the items found by the systematic crawl, implying that nearly all of the artifacts on the surface were found in the first survey pass. The results of the crawling survey are considered a close estimate of the total number of artifacts on the surface at the time of the survey and can be used to evaluate the accuracy of coarser-grained coverage of the same area.

In comparing the walking to the crawling surveys (Table 2), only the 130 sq m area that both methods cover was included (subplots 1 to 10, A, B, and C). The systematic and nonsystematic discoveries are combined for each subplot. Comparison of the techniques indicates that the crawling survey increased the total number of chipped stone artifacts discovered per square meter by 362% (TABLE 2). The spe-
Table 2. Comparison of total and average chipped stone (CS) discovered by walking and crawling surveys, with same area covered (130 sq m). Plot locations are shown in Figure 5.

<table>
<thead>
<tr>
<th>Modified Whittaker plot</th>
<th>Walking survey</th>
<th>Crawling survey</th>
<th>Walk vs. crawl</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total CS Av/ sq m</td>
<td>Total CS Av/ sq m</td>
<td>Net increase % in/rote</td>
</tr>
<tr>
<td>NRTJ</td>
<td>65 0.50</td>
<td>244 1.88</td>
<td>179 375</td>
</tr>
<tr>
<td>WRTIN</td>
<td>6 0.05</td>
<td>10 0.08</td>
<td>4 167</td>
</tr>
<tr>
<td>WJWTU</td>
<td>0 0</td>
<td>4 0.03</td>
<td>4</td>
</tr>
<tr>
<td>CC01</td>
<td>2 0.02</td>
<td>7 0.05</td>
<td>5 350</td>
</tr>
<tr>
<td>WH01</td>
<td>0 0</td>
<td>0 0.00</td>
<td>0</td>
</tr>
<tr>
<td>33N6</td>
<td>1 0.01</td>
<td>3 0.02</td>
<td>2 300</td>
</tr>
<tr>
<td>N6BS</td>
<td>0 0</td>
<td>0 0.00</td>
<td>0</td>
</tr>
<tr>
<td>WTRI</td>
<td>0 0</td>
<td>0 0.00</td>
<td>0</td>
</tr>
<tr>
<td>WTRW</td>
<td>0 0</td>
<td>0 0.00</td>
<td>0</td>
</tr>
<tr>
<td>TTO1</td>
<td>0 0</td>
<td>0 0.00</td>
<td>0</td>
</tr>
<tr>
<td>Total, mean</td>
<td>74 0.06</td>
<td>268 0.21</td>
<td>194 362</td>
</tr>
</tbody>
</table>

* Note that four artifacts were found crawling and zero were found walking. This is a significant increase but a percentage cannot be calculated.

The specific value of 362% is not meant to be generalized to other contexts, but it is clear that pedestrian surveys of any transect width are missing many artifacts in most settings. The magnitude of this increase was greater than many archaeologists (ourselves included) might have assumed and the implications need to be considered.

The size of artifacts alone does not explain the increase in items found crawling. For the walking survey, the average maximum length of artifacts discovered during the systematic pass is 17.3 mm, but the nonsystematic finds were only 12.7 mm long on average. Thus, the initial survey pass by the walking crew favored larger items. The artifacts found by the crawling survey averaged 13.8 mm long, which is not significantly smaller than the combined average of 14.6 mm for all discoveries made by the walking survey. Although non-systematic coverage recovered small artifacts, the crawl survey led to a more accurate portrayal of the surface record.

**Linking Nested Samples**

Regression analysis of each survey demonstrates that the ability to predict the properties of the sub-surface archaeological record is much greater for the crawling survey than the walking survey. The crawling survey could explain 72% of the variance in the top 10 cm whereas the walking survey explained only 24%. Interestingly, no combination of ground surface visibility, vegetation height, or plant biomass improved the amount of explained variability. When archaeologists are deciding where to place test excavation units, often the only clues available are depositional setting and surface artifact density. This provides a starting point for understanding the degree to which the surface record “predicts” the subsurface record in different settings. The surface and near-surface records are not separate entities and differences between them, in some instances, may be more a function of sample accuracy than the nature of archaeological deposits.

While this suggests that high resolution samples can predict some of the variation in the near-surface record, the archaeological reality is that in situations such as pipeline surveys, archaeologists must often attempt to predict the nature of deposits far below the surface based on surface properties. Such situations are difficult because of the expense of conducting deep test excavations combined with the problematic reliance on visible distributions as indicators of materials far below the surface. We do not suggest that crawling surveys have the ability to predict artifact densities far below the surface, and ideally evaluations of buried deposits must be augmented with exposed vertical faces or test excavations and trenches. Furthermore, these relationships will likely change significantly with larger sample sizes and especially in different depositional contexts and high surface density settings. The large number of zeroes in the dataset indicates that NRTJ, the high surface density plot, had an especially strong influence on the outcomes of the regressions. In spite of this, expanding our understanding of the factors that condition variance in surface and subsurface samples will be useful for property-based evaluations and for comparing samples across contexts.

**Discussion**

Currently, there is little consensus on the ideal transect spacing for pedestrian surveys. In many instances this may

---

2. If the variables are log-transformed, the crawling survey explains 54% of the variance in the near-surface record and the walking survey explains 18%.
be a project-specific decision dependent on variables such as funding, time constraints, and the goals of the project. When different methods are used in different projects, however, the results are difficult to compare. For perspective, Schiffer and Wells (1982: 353) present a table containing crew spacing and other details of 12 surveys that took place in the Southwest United States between 1974 and 1980. During this time, transect widths varied from 4 to 50 m but averaged ca. 25 m. Current archaeological practice still exhibits this range of variability. Surveys that claim complete coverage of a region generally use wide transects, but are often geared towards documenting highly visible features (compared to chipped stone) such as buried structures, temples, or cities. While all major settlements may indeed be found by such surveys, the notion of 100% coverage requires that a substantial portion of the surface record be written off as insignificant. To return to the screen size comparison, claims of 100% coverage are akin to using a very large screen size in an excavation and asserting that everything of importance is larger than the gaps in the mesh.

Compared to the norms of current practice, our 70 cm transect width is absurdly narrow and should yield highly accurate results in spite of the limited spatial coverage. The results of the walking survey with 70 cm spacing, however, explained only a fraction of the variance in the subsurface records and a tremendous amount of surface chipped stone was overlooked. Wider transect spacing progressively lessens the ability to address basic parameters of the target population. This suggests that an optimal strategy may involve multiple phases of survey, one of which should be highly consistent and precise, while another would focus on large coarse-grained samples. The need to protect sites from land disturbance and the fact that researchers will never know where all cultural properties are located requires the use of conventional pedestrian transects. In the case of land management districts which must regularly contract for surveys, implementing just a few Modified-Whittaker plots in each survey would over the course of time lead to a far better understanding of the regional record. The methods outlined in this article would be especially useful if applied and experimented with in a wider range of contexts.

Area and Intensity

Two primary causes of systematic surveys failing to discover cultural materials can be identified. One is the sacrifice of area: the smaller the spatial extent of the regional sample, the more material within that region will escape discovery. The second is the sacrifice of intensity: the lower the observational intensity, or grain, used within the surveyed area, the less material within that area will be observed. Sampling schemes employed in survey projects should consider both of these sacrifices. This is a difficult problem because both sacrifices appear to be inevitable, while neither is acceptable. Which is the "better" way to not find artifacts: to not look in enough places, or to not look closely enough? While this is a trade-off (Schiffer and Wells 1982: 347), nested-intensity surveys and multi-scale survey designs introduce ways to make quantitative sense out of these two respective sacrifices.

In considering the sacrifice of area, the problems are the placement of the plots on the landscape and the spatial extent of the sample units. Some of our plots demonstrate that if only small areas are crawled, the area sacrificed could lead to a complete absence of discovery. The increase in items found by the crawling survey demonstrates that the sacrifice of intensity can be significant. When walking surveys fail to find materials, this may be because nothing was present or because nothing was identified. An area crawled with nothing discovered is not a wasted effort, because the confidence of knowing that no artifacts were visible at the time of the survey is in itself valuable. Without conducting a limited amount of survey at this intensity, the simple question of "how much material is present within the surveyed area?" cannot be addressed. This point can also be made in statistical terms. A sample is representative if its values approximate those of the population. In the case of standard archaeological surface samples, basic properties such as mean and median artifact density cannot be calculated reliably (Van de Velde 2001: 25).

Archaeological surveys frequently claim to have achieved complete coverage of a region. In this sense, the intent is to demonstrate that the entire region has been observed so that the researcher does not have to infer regional properties based on limited samples. Coarse-grained coverage of large regions is necessary when the goal is identifying relationships between settlements is, and such surveys are often focused on the discovery of highly visible features. The claim that surveys result in 100% coverage is overly optimistic, however. Interpretations and management decisions based on pedestrian samples should be tempered by the properties of survey data. In any setting, archaeologists need to be more explicitly aware of the inferences that can be reliably made from a sample given the nature of the coverage. Again, the techniques presented here may be useful for reaching more accurate evaluations of the nature of archaeological survey data that can be used to complement coarse-grained coverage.

Other approaches to surface survey have emphasized collection strategies, such as Van de Velde's (2001) point-sampling approach that collects artifacts in 2 m diameter
circles on a systematic grid. Artifact collection is not appropriate in all circumstances, but Van de Velde applies systematic collection to the problem of obtaining accurate samples. When archaeologists must conduct research as well as preserve and manage the record they study, it makes little sense to destroy the record while also managing it. Obviously, some artifacts require analysis that cannot be conducted in the field and must be collected. In other cases, the record is in danger of destruction and collection may be necessary. Van de Velde's systemic collection strategy is a novel approach, but isolated 2 m circles will not reveal the distributional properties of a regional record nor can they be used to evaluate methodological accuracy or stability in surface distributions. The Modified-Whittaker plot's advantages over such approaches include the expanded spatial sample, improving the probability of recording isolated and rare finds, and the ability to investigate the record at multiple scales (Stohlgren et al. 1997).

The General Utility of Multi-Scale Sampling

Although the ONG Survey Project area is a hunter-gatherer record composed almost entirely of chipped stone artifacts, the utility of the techniques presented here is applicable in a variety of archaeological contexts. Archaeologists studying complex societies often deal with extensive scatters of ceramic sherds. Attempts at making inferences from these surface scatters and their associated features developed into the conceptual cornerstone for virtually all contemporary survey projects (e.g., Willey 1953). Following this tradition, settlement pattern analysis led to a variety of high-level interpretations regarding prehistoric population densities, movements, and cultural changes that were largely based on the analysis of surface scatters (e.g., Sanders, Parsons, and Santley 1979). These estimates are based on calculating the number of different ceramic types, sherd density, and site size. Multi-scale techniques could be valuable in such situations because they allow for frequencies of different ceramic types to be investigated as spatial scale increases. Application of the Modified-Whittaker plot could enhance and refine the data gathered during the pioneering survey projects conducted in these regions and improve comparisons between sites of different types or between regions.

It might be argued that this method is useful for low artifact densities in arid landscapes but could not be used in regions of either extremely high artifact loads or high vegetation cover. Certainly, with the restraints of time and money, piece-plottting each individual artifact may not be viable in all situations. The advantages of multi-scale sampling are not dependent on accurate artifact-level provenience, however. Archaeologists frequently work in areas where the ground surface is blanketed in artifacts. How should such populations be sampled? Very dense artifact scatters can also be sampled with intensive strategies and may especially benefit from multi-scale techniques. The multi-scale framework of the Modified-Whittaker plot can be used for projecting the properties of the sample to the complete population.

To apply the Modified-Whittaker plot to the sampling of dense surface records, the abundance and diversity of each artifact type would be totaled for each of the subplots and the K plot. These data can then be plotted with the spatial scales of 1, 10, 100, and 1000 sq m on the x-axis with either the number of observed artifacts or the number of artifact types on the y-axis. By fitting a function to these points, the area-density relationship can be projected to make quantitative estimates for how the sample should behave at scales larger than can be feasibly observed. This function could be used to estimate the number of individuals for a certain type or to project the diversity of the assemblage in a manner analogous to the plant species-area curve (Begon, Harper, and Townsend 1996: 861–871; Mueller-Dombois and Ellenberg 1974: 51). As area increases, so will the diversity of the sample. This is a function of the larger sample size and the higher probability of finding rare specimens in a larger area, but the exact nature of the relationship will depend on the heterogeneity of the population (Ebert 1992: 213–244). Artifact diversity should also increase with area and analyzing distributions in an area-diversity perspective may be a useful avenue for comparing typologically diverse archaeological landscapes.

Conclusion

In this paper we have attempted to show how property-based investigations of the archaeological record can be more fruitfully implemented by modifying a few aspects of traditional survey practice. By combining this approach with existing norms, the context of the items discovered is enhanced by incorporating a more accurate understanding of the nature of the population that is being sampled. This is useful for pragmatic concerns such as the effect of transect spacing on artifact density, and can also be used for monitoring change over time on an archaeological landscape. Understanding how processes such as large herbivore grazing influence the distribution of archaeological materials is an important management concern (e.g., Minard 2003) and is also necessary for archaeological inter-

3. Most computer programs that can form scatter plots such as Excel can fit functions to the data distribution and provide the resulting equation of the function. A power function is the most appropriate for applications like this.
interpretations of spatial relationships among artifacts and features. For instance, we found that single seasons of grazing can seriously disturb artifact distributions and the approach we used to monitor them can facilitate the development of a framework for estimating long-term effects. We also found that using high-resolution coverage can lead to important revelations, such as the magnitude of artifact loss given certain techniques and the degree to which the surface record can change over short time scales. These processes need to be investigated with an experimental approach and the Modified-Whittaker plot provides an ideal framework for doing so.

A number of conclusions could be made. Perhaps archaeologists should resign themselves to the fact that surface samples gleaned by survey will not match the properties of the surface record, and we should proceed as normal, simply regarding this as a cautionary tale. On the other hand, many archaeologists would argue for the need to continually seek new ways of improving management and research techniques in the attempt to extract behavioral information from surface materials. We prefer the latter option because a failure to understand the properties of the distributions captured by a survey will lead to a failure to understand the processes behind them.

Acknowledgments

The participants in the Colorado State University field school from 1999 to 2001 were enthusiastic, dedicated, and most of all, great to work with. We are indebted to them for making this research possible. We are also thankful for volunteers from previous field schools who returned to take part in the research. LuAnn Wandsnider provided useful commentary and an excellent foundation for our experiments with method and discovery. Judson Finley and Dave Rapson made valuable contributions to early drafts of this paper. Three anonymous reviewers offered especially valuable insights and editorial suggestions.

Oscar Burger is a graduate student at the University of New Mexico interested in landscape archaeology and understanding the effect of human behavior on the environment. This manuscript arose from his Master's research, completed in the spring of 2002 at Colorado State University. Mailing Address: Department of Anthropology, University of New Mexico, Albuquerque, NM 87131-1086. E-mail: oskar@ unm.edu

Lawrence C. Todd organized the ONG research project and is interested in taphonomic research across spatial scales and the development of integrated multi-disciplinary research and management perspectives.

Paul Burnett is a graduate student at Colorado State University and is developing applications of these techniques for research in high-altitude settings.

Tomas J. Stohlgren is a research scientist for the U. S. Geological Survey in Fort Collins, Colorado. He studies plant biodiversity and sampling designs in North American ecosystems.

Doug Stephens is a Forest Service archaeologist and oversees the research in the Oglala National Grasslands. He is interested in integrating field techniques that meet both management and research requirements.


Mueller-Dombois, Dieter, and Heinz Ellenberg  

Orton, Clive  

Sanders, William T., J. R. Parsons, and R. S. Sandley  

Schiffer, Michael B., and Susan J. Wells  

Schiffer, Michael B., A. P. Sullivan, and T. C. Klinger  

Short, Michael J.  

Stohlgren, Tomas J., K. A. Bull, and Y. Otsuki  

Stohlgren, Tomas J., M. B. Falkner, and L. D. Schell  

Stohlgren, Tomas J., G. W. Chong, M. A. Kalkhan, and L. D. Schell  

Thomas, David H.  

Todd, Lawrence C.  

Van de Velde, Pieter  

Wandsnider, LuAnn, and Eileen Camilli  

Willey, Gordon R.  