Analysis of Archaeological Sampling Methods Using the Complete Surface Data from the Pirque Alto Site in Cochabamba, Bolivia

Elizabeth Green

Faculty Sponsor: Timothy McAndrews, Department of Sociology/Archaeology

ABSTRACT
Sampling is an extremely important aspect of archaeological fieldwork, and it can significantly influence interpretation; therefore, the effects of these processes must be understood and controlled. Various experimental sampling methods are tested against the surface recovery data collected from the Pirque Alto site in Cochabamba, Bolivia. The project addresses the following questions: do results from the sampling methods reflect the total surface recovery; what method gives the most accurate results for this site and why; and at what percentage of quadrat (unit) samples taken per method do the results show validity. Finally, in what way are the methods biased and how can these biases be addressed? The project has yielded information about how various methods would or would not have affected the interpretation of the Pirque Alto site, and the results were then applied to a broader study and understanding of sampling methods.

INTRODUCTION
Sampling, a less visible but just as essential tool in archaeology as the trowel, has been used since the start of the discipline. “Archaeologists have always sampled… archaeologists have been choosing, among alternatives, where to look for archaeological sites, which sites to excavate, and where on selected sites to dig” (Hole 1980:217). However, despite the fact that archaeologists have always sampled, there seems to be a disproportionately small effort put forth to truly understand and test the methods so heavily relied upon. Sampling exists and is used for many reasons, including: financial, labor and time restraints and the logistical necessity of controlling the sheer volume of physical materials and non-physical data generated by the process of archaeology.

The processes of sampling have far-reaching consequences and influences. For instance, a sampling of units on the surface of a site can directly affect an archaeologist’s perception of the site (its function, chronology, scope, etc), as well as directly affecting the direction of the research questions posed in the future. It is important to know that these critical procedures and methodologies are the most conclusive possible so that their influence on the excavation and interpretation of sites is as constructive as possible. “Given that sampling is essential, we need to ensure that it is carried out in a way that enables us to make best use of the data that it provides” (Orton 2000:8).

Due to this, it is important that experiments with comparable results be conducted in order to gain a better understanding of the act of sampling and its effects. The purpose of this specific project is to do just that — to act as a testing mechanism for common sampling procedures. The data used in this endeavor comes from the work done at the site of Pirque Alto, located in the Department of Cochabamba, Bolivia. Details of this site will be discussed in the background section to follow; however, it is important to note that the work conducted at this site included a surface collection of ceramics and lithics of the entire site. Ceramics bigger than a nickel and all of those featuring paint were collected.

The collected ceramics data act as a control group in testing various sampling methods with the goal of answering the following questions:

1. Do results from the sampling methods equal or come close to the spatial (ex. identified ceramic clusters) and numeric results indicated from surface collection?
2. What method gives the most spatially and numerically accurate results for this site and why?
3. At what percentage of unit samples taken per method do the results show the greatest spatial and numerical accuracy, (10%, 30%, or 50%)?
4. In what way are the methods of sampling biased, and how can these biases be addressed or corrected?

The ultimate goal of this project is to determine the validity and precision of common archaeological sampling methods and identify the most accurate methods. Because the data being used as the control group comes from a real site, looking at how the sampling methods would have affected its interpretation help to present the gravity of influence from sampling.

BACKGROUND

The Site of Pirque Alto

During the 2005 field season, archaeological fieldwork was conducted by several students in the UW-L Archaeological Studies program under the supervision of Dr. Timothy McAndrews at Pirque Alto, in the Department of Cochabamba, Bolivia. One goal of this project was to investigate the influence of the Tiwanaku culture in the area through the cultural remains present at the site, such as ceramics, stone artifacts, and features. Tiwanaku was a powerful polity operating out of the southern part of the Lake Titicaca Basin that influenced a broad region throughout the central Andes from A.D. 500 to A.D. 1150 (Janusek 2004:xvii). Now viewed as “the first true state in the south-central Andes,” the Tiwanaku culture is typified by stratified societal levels, monumental architecture, and sophisticated material culture (Young-Sanchez 2004:17-18). Their material culture is unique, especially the ceramics which are “extremely distinctive and rather easy to recognize both in terms of slip and paint, and in diameter of base and rim,” which allows for easier identification (Burkholder 1997: 130). The nature of the spread of the Tiwanaku is a topic currently under fierce debate, and as such this site could help shed light on the presence of the Tiwanaku in the Cochabamba Valley area.

Through the surface collection of 406 5 x 5 meter units, a total of over 13,200 diagnostic ceramics (rims, bases, handles, painted pieces), and over 52,000 non-diagnostic (those pieces bigger than a nickel not belonging to the diagnostic category) were collected and subsequently analyzed. Early conclusions regarding this site indicated that it was a multi-component site occupied from the Formative Period though Inca times. This site is located on a bluff

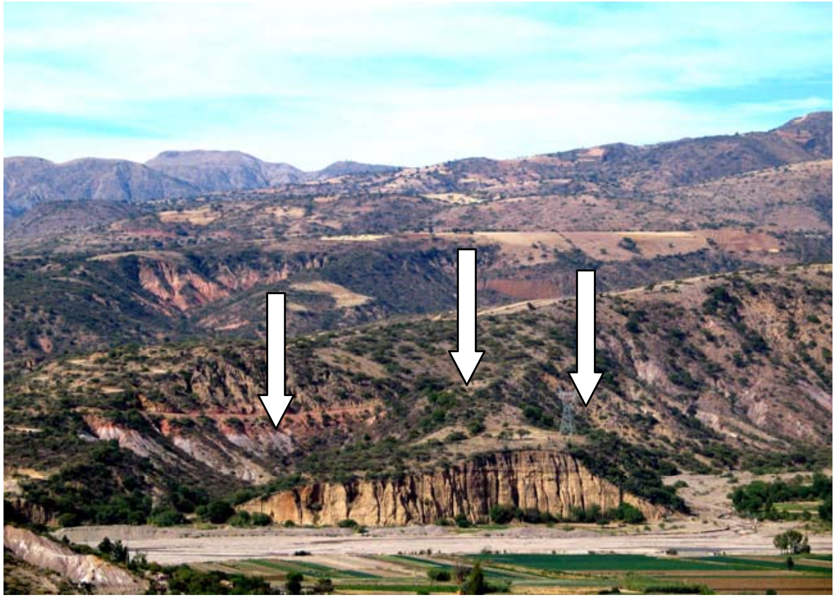

Figure 1. Site of Pirque Alto as viewed from opposite valley wall

(see Figure 1) overlooking the Rio Tapacari and exists in a natural travel corridor leading to the Altiplano and Titicaca Basin. Since this site, in terms of location and artifact assemblage, can offer important addition to the debate concerning Tiwanaku presence in the Cochabamba Valley, it serves as an excellent case of how sampling would and could affect site interpretation and how in turn that can affect the archaeology of an entire area (McAndrews 2006).

History of Sampling

The history of sampling in archaeology is an interesting one, and the current debates within the subject showcase the breadth of thought concerning its importance and its influence, both on the process of archaeology and on archaeologists themselves. One of the few comprehensive histories of sampling in archaeology is contained in *Sampling in Archaeology,* by Clive Orton (2000). In this book, Orton discusses the changes in the actual sampling themes throughout time but also talks about the changes in archaeologists’ attitudes towards sampling.

At the turn of the twentieth century, there was more attention directed towards sampling; however, without the later developments that statistics would bring, “it would not be reasonable to expect more than an informal approach” (Orton 2000:112). Preliminary attempts at sampling were carried out in a rather intuitive way. For instance, if there appeared to be a large amount of sherds, you took what you believed to be “representative.” It was not until the developments of the 1940s that sampling procedures became more concrete and statistically representative. The developments in statistical theory came about as a result of the need for quality control in response to industrial growth but eventually became incorporated into the sampling of shell middens. This helped to cut down on the tedium of going through mountains of low grade data. At first glance, the beginning of sampling in archaeology would seem to suggest a rather haphazard approach designed to make archaeological evaluation easier and less tedious.

It was during the 1950s and 1960s that more formal sampling methods, which Orton describes as “samples selected from well-defined populations according to rigorous statistical procedures,” became more commonplace (Orton 2000:1-2). Vescelius was the first to discuss the use of probabilistic sampling, and he directed his attention towards finding a way to estimate artifact proportions compared to all others within a site (Vescelius 1960:457-470). However, “no other single article has altered more radically the recent course of archaeology than Lewis Binford’s 1964 consideration of research design” (Hole 1980:218).

Lewis Binford contributed greatly to the field of archaeology, especially in relation to field methodology and research design, which included sampling. Although Binford’s regard to sampling was directed regionally rather than at a site level, he instituted tremendous changes in the way archaeology was carried out. Orton describes there being three major changes advocated by Binford: “the promotion of the region to the position of primary unit of archaeological research, the explicit design of archaeological research programs, and the explicit use of sampling theory as the way of linking the two” (Orton 2000:4).

Though Binford’s contributions to the field of archaeology cannot be underestimated, such an enforcement and indoctrination of sampling methods caused a severe backlash that emerged in the 1970s. Sampling appeared as an evil necessity that one would never escape and as a compromise with the constraints of real life. This feeling increased with the development of sieving methods at the feature level, which created even more data which required sampling from (Orton 2000:4). Part of this backlash was to criticize sampling itself rather than how it was being carried out, as though sampling was something impressed upon archaeology by fields that had nothing to do with archaeology.

Though the 1970s are characterized by a backlash against sampling, they were also a time of reflection on the progress of sampling due to the publication of the San Francisco Symposium (which was the first formal gathering to discuss sampling) and Kent Flannery’s *Early Mesoamerican Village* (Flannery 1970). James Mueller, one of the big names in sampling theory, describes this reactionary period as extremely varied. “The responses and reactions have been varied – categorical rejection, blind acceptance, and skeptical positivism,” the last which seems to dominate in the 1970s (Mueller 1975: xi).

This evaluation of the progress of sampling had the curious result of bringing viewpoints regarding it back full circle to a more intuitive process. “By the 1980s in the USA and the 1990s in the UK, sampling had become firmly entrenched in the methodology of the ‘contract’ wing of archaeology” and had become something almost instinctive among archaeologists (Orton 2000:6). Though sampling has again become something rather intrinsic to archaeology, “[it] is still on the agenda even if it is not discussed as openly or explicitly as it was, say, in the 1970s and 1980s” (Orton 2000: 207).

Debates within Sampling

Due to the importance of sampling, there exists a great deal of controversy throughout the literature. The debates range from the general importance of sampling (Rootenberg 1964) to the proper methods while conducting sampling (Schiffer, Sullivan and Klinger 1978) to the correct math and proportions (Plog and Hegmon 1993, Nance 1981). According to Bonnie Laird Hole, “Despite the widespread availability of understandable presentations of the theory of sampling and the powerful potential of its appropriate application in archaeological research, its contributions to archaeological knowledge have been disappointing” (Hole 1980:217).

Articles run the gamut from math-entrenched to highly theoretical and behavioral. It is interesting to note that within most of the literature there exists a great deal of disagreement expressed through journal publications. At times some archaeologists would rather attack one another through publication than to correspond directly. Though opinions are varied, it is important to mind the theoretical milieu from which the disagreements arise.

A great debate going on in sampling right now is the degree to which post-depositional factors contribute to the bias of a sample. Many in the Processual archaeology group bypassed looking at this question by examining things which were the least affected. The Cultural History school of thought also tended to ignore the issue. (Orton 2000:43). This debate continues today, especially in regards to issues like the value of the number of identified specimens over the minimum number of individuals.

A debate more relevant to this project is the degree to which surface remains reflect the subsurface, and "is a key issue for the practice of archaeological survey, at both the regional and the site level. Opinions, both as expressed and as implicit as fieldwork practices, seem to vary widely" (Orton 2000:57). For the site of Pirque Alto, the clusters of certain identified phases are so distinct and identifiable that they are assumed to represent a reliable indication of the subsurface.

Obviously, these are hardly the only ongoing debates within sampling, and more research is being conducted in order to present specific debates concerning individual sampling methods; i.e. random vs. judgmental, etc. Some of these disagreements will be explored later in the paper as they relate to the results of the project.

Descriptions of the Sampling Methods used in this Project

Below are basic descriptions of the types of sampling methods relied upon in this project. Information concerning the experimental procedures will be discussed in the methodology section of this paper. Lastly, more specific pros and cons for each method will be discussed in the results section as they correlate directly. Archaeological sampling methods can be roughly divided into two categories: probabilistic and non-probabilistic. Archaeology employs probabilistic sampling methods as an "attempt to improve the probability that generalizations from the sample will be correct" (Renfrew and Bahn 2000: 76). Charles Redman describes probabilistic sampling methods as "good for estimated total population values for common artifacts or features… probability sampling is poor for locating rare feature or artifacts, dealing with clustered distributions, or illuminating contiguous spatial patterns" (Redman 1987: 251). Non-probabilistic methods, described subsequently, are generally looked upon less favorably.

The most basic probabilistic sampling method is a simple random sample, in which numbers are assigned to the quadrats and those to be sampled are chosen through a random number chart till a predetermined number or percentage of units is sampled. "It implies that an equal probability of selection is assigned to each unit of the frame at the time of sample selection" (Binford 1964:141).

Systematic random sampling involves the sectioning out of a site grid into groups of a predetermined number of quadrats, and then randomly selecting quadrats within each section (Drennan 1996: 246). The general advantage of this method is that it ensures there will be broader coverage across the entire site; "Systematic sampling ensures an equal dispersion of sample units" (Binford 1964:153).

Next there is stratified random sampling, where the site is divided into its natural zones (or strata) and like before, random numbers are chosen. However, in this method the squares are proportional to the natural zones; 85% forest equals 85% of the samples are taken from that area (Renfrew 2000: 77).

Judgmental sampling, which is not representative of the whole, is a non-probabilistic form of sampling and is associated with similar methods including haphazard sampling, grab sampling, and purposive sampling. "Non-probabilistic sampling is regarded as intuitive, inductive, and unstated" (Hole 1980: 219). All of these terms represent a method referring to "explicit or implicit application of a variety of nonrandom selection criteria" (Drennan 1996:88). In judgmental sampling, samples are "selected by looking over the range of elements in a population and specifically deciding to include certain elements in the sample and exclude others" (Drennan 1996:88). Obviously, this method has a large degree of bias. It was important to include due to the current instinctive feeling of archaeological sampling.

METHODOLOGY

The first step of this process was to become familiar with the literature concerning sampling. This was an important step because it helped to frame the concerns and debates within sampling and helped to provide a general understanding of the subject. Another familiarization step was to become comfortable with the computer programs necessary for this project, including Microsoft Access and Excel, and Surfer by Golden Software, a map generating program. After sufficient skills with these programs had been developed the sampling experiments could be carried out.

Sampling experiments were carried out at three percent intervals (10%, 30%, and 50%) with five trials each. Originally, these experiments were to be carried out for the total ceramics data and for the identified Tiwanaku phase ceramics. However, a different approach has been considered and decided on instead. Each of the trials will take into account the total ceramics data, yet the data will show the percent levels for three identified ceramic phases: Formative, Early Intermediate, and Tiwanaku. These percents will be compared to the total percents for each of the ceramic phases. For example, if a 10% random sample shows 13% of sherds belonging to the Tiwanaku phase ceramics, this will be compared to the percent of Tiwanaku phase ceramics for the entire site.

The sampling experiments focus on four types of sampling commonly relied upon by archaeologists: random sampling, systematic sampling, stratified sampling, and judgmental sampling. For the random sampling experiments, Microsoft Access was used to randomly sample units according to the percent level specified. For the systematic sample, the site is divided into as many 25 x 25 meter groups as possible. The units within these groups are then selected randomly according to percent. See Figure 2 for systematic group distributions.

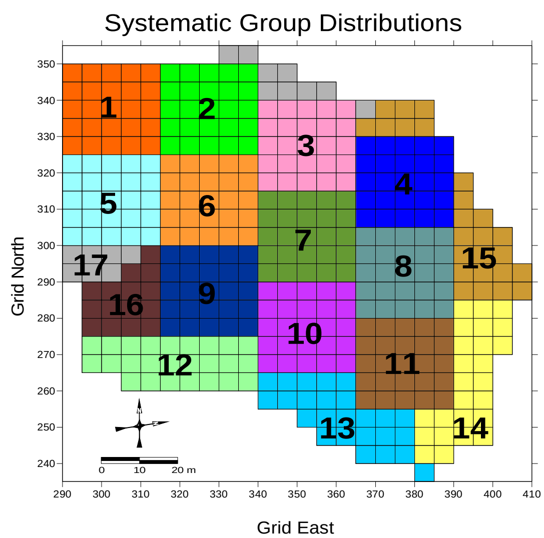

Figure 2. Systematic Group Distributions

For the stratified random sample, the site is broken up into three areas related to their relative elevation: highest part of the sites, middle and lowest. Grids are then selected randomly according to percent level. See Figure 3 for the stratified group distributions.

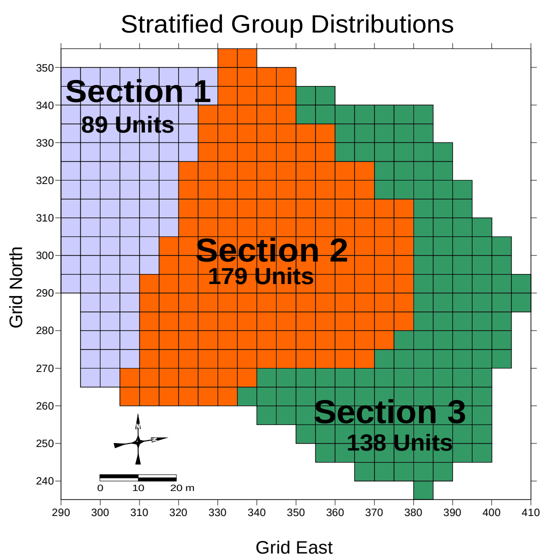

Figure 3. Stratified Group Distributions

Since the judgmental sample is a non-probabilistic method, it was carried out differently than the previous methods. Three students in the Senior Thesis class (ARC 499) who did not attend the 2005 field school at Pirque Alto were given a site map with general information such as the physical layout of the site, areas of higher artifact concentration (like one would see on a pedestrian survey). They were then asked to indicate which areas based on the information they had been given would be their choice for sampling. Size of the sampled areas was equal to the percent levels used for the other three sampling methods. For example, the first student was given enough units to equal a 10% sample of the site to lay out as they see fit; for the second student, 30% and so on. Though this method differs significantly and will not be able to be compared in exactly the same way, it is an important avenue to explore.

After the experiments were carried out, the information was plotted onto a site map using Surfer by Golden Software. This allows for a clear illustration of the differences between the sampling methods and their visual results. A series of descriptive statistics (mean and standard deviation) were generated and will be used in comparison to the visual results. T-tests were also carried out to determine if there was a statistically significant difference between the samples and the full dataset in terms of the relative proportions of different periods of ceramics. It was determined that there was not a statistically significant difference; meaning all samples may be considered numerically representative of the whole.

Armed with both the visual representations of the sampling experiments created using Surfer and the statistics generated from the samples, a comprehensive comparison of the methods will then occur. The relative success or failure of the various sampling methods will be viewed in regards to their implications for interpretations of the site and then applied on a larger scale to view the effect of sampling on the field of archaeology.

RESULTS

In addition to the main goal of determining the validity and precision of common archaeological sampling methods, this thesis also sought to address four specific questions:

1. Do results from the sampling methods equal or come close to the spatial (ex. identified ceramic clusters) and numeric results indicated from surface collection?
2. What method gives the most spatially and numerically accurate results for this site and why?
3. At what percentage of unit samples taken per method do the results show the greatest spatial and numerical accuracy, (10%, 30%, or 50%)?
4. In what way are the methods of sampling biased, and how can these biases be addressed or corrected?

These questions will be addressed throughout this portion of the paper, which is itself broken up into three different sections. In the first section, results in terms of the individual methods (random, systematic, etc) will be discussed. This will be followed by a section describing the results on each percent interval, including the spatial representations of each sample method and the statistics and ceramic phase percentages. This is done to illustrate the benefits and problems associated with each method and provide more detailed information at the sample size level. General comments about the various methods and their results will comprise the last section. On the following page, Figure 4 shows the overall distributions of all ceramics over the entire site for reference. Maps for the Formative, Early Intermediate, and Tiwanaku Ceramics can be found at the end of the document.

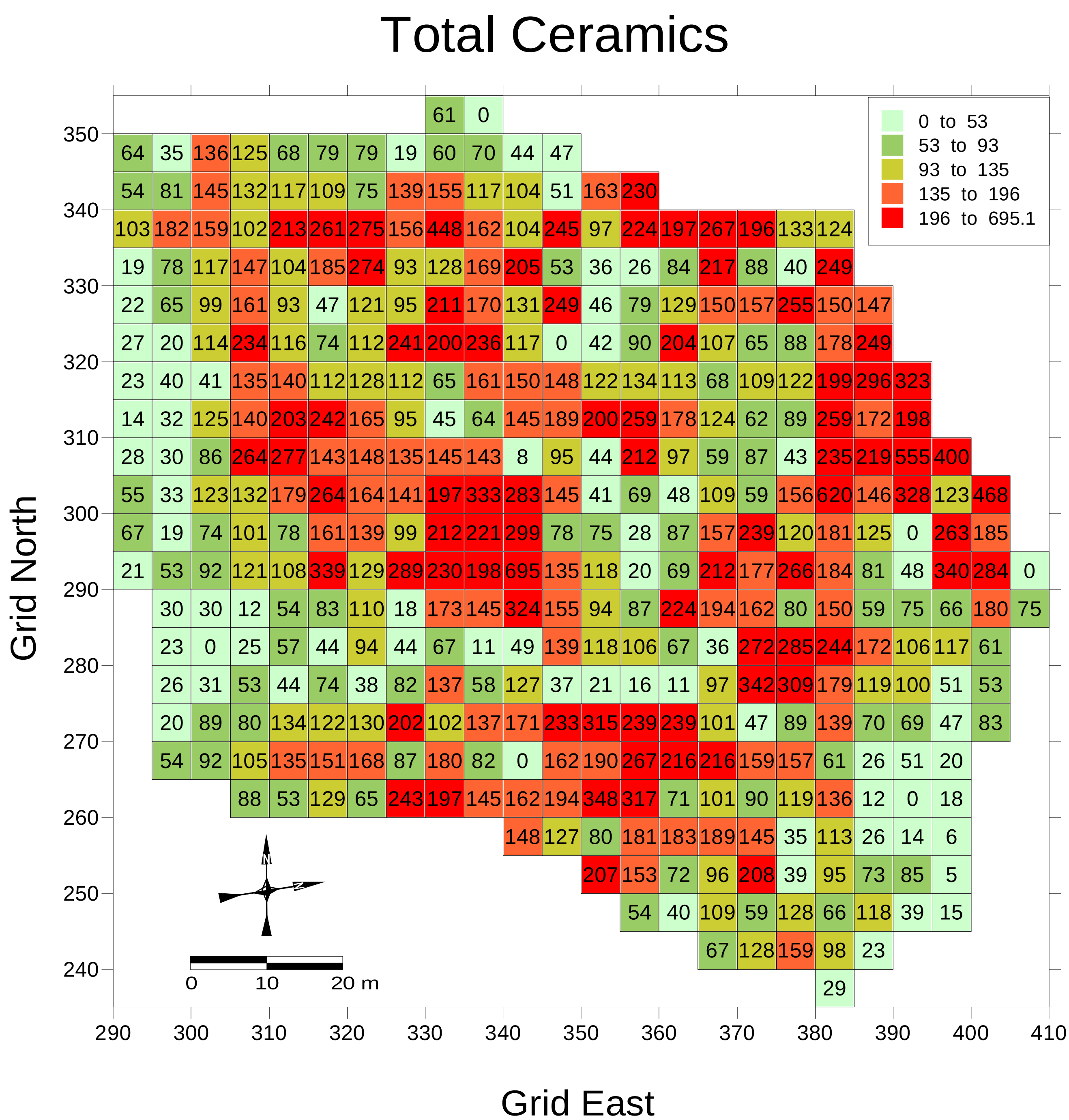

Figure 4. Total ceramics distribution

Individual Sampling Method Results

Random Sampling. Without a doubt, random sampling produced the most varied results, most notably within the spatial distributions. At odds with this is that the statistical and phase percentage results gave a representative sample. The circled areas on the maps indicate areas of high frequency identified ceramic phase clusters that would have been missed by the sample. For the 10% random sample (see Figure 5), the circled upper right-hand and lower section contain the two largest clusters of Tiwanaku ceramics on the site. As one of the goals of the Pirque Alto project was to gather information about the Tiwanaku influence at the site, missing these clusters would have drastically affected our understanding of the site.

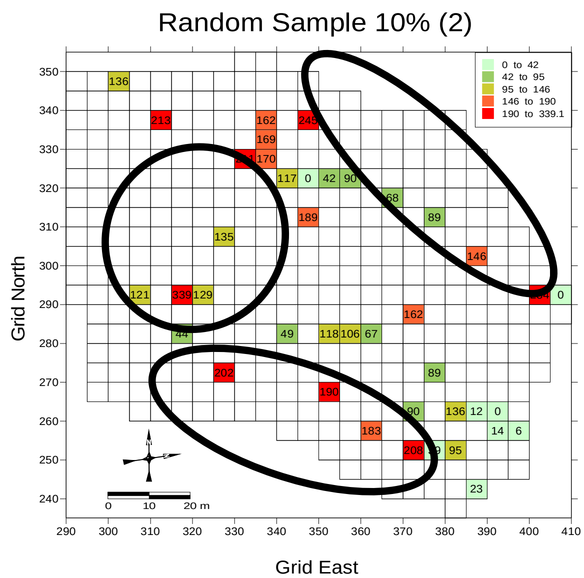

Figure 5. Random Sample 10% (Trial 2)

The above map was generated from the 10% sample level, and it can be seen that is performs poorly in locating the major clusters of identified ceramics. For this sample type, this trend continues through the 30% samples. Figure 6 shows that even at three times the previous sample size a large chunk of the site was not sampled.

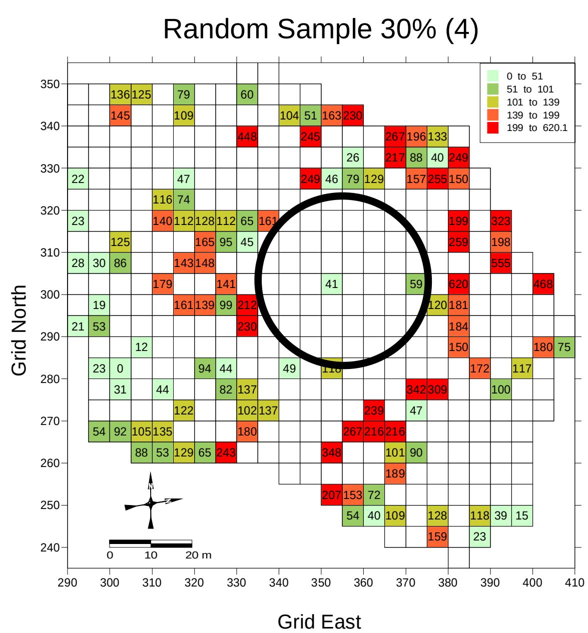

Figure 6. Random Sample 30% (Trial 4)

For the random sample, the most reliable sample size level was 50%, (see Figure 7).

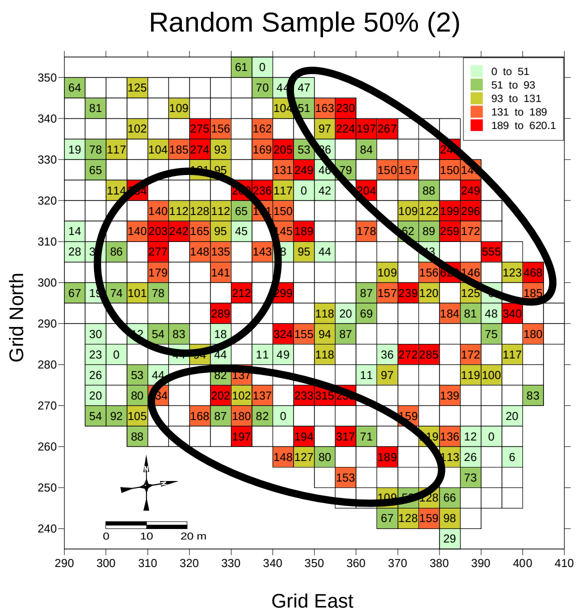

Figure 7. Random Sample 50% (Trial 2)

Table 1 shows the numeric results for all three maps shown above. Despite the large spatial differences between all three intervals, the random samples produced the most reliable and representative numeric results compared to both the systematic and the stratified samples. However, when viewed in conjunction with their spatial distributions, random sampling comes up quite short. For a statistician random sampling would be the most ideal, but for the archaeologist who deals spatially and numerically, random sampling has little to offer. “In an archaeological context, simple random sampling will not work” (Rootenberg 1964:182).

Table 1. Random Sample Numeric Results

	Whole Site	RS 10% (Trial 2)	RS 30% (Trial 4)	RS 50% (Trial 2)
Number of Units Sampled	406	41	121	203
Mean ceramics per unit	128.81	119.22	138.66	126.98
Formative Phase %	4.46%	3.85%	5.26%	3.99%
Early Intermediate Phase %	0.90%	0.94%	0.84%	0.96%
Tiwanaku Phase %	18.99%	19.33%	18.47%	18.70%
Diagnostic Ceramics %	25.24%	24.96%	25.30%	24.54%
Non-Diagnostic Ceramics %	74.76%	75.04%	74.70%	75.46%

Systematic Random Sampling. The systematic random sampling method provided the most consistent results of all three probabilistic sampling methods spatially and numerically, and therefore it was difficult to choose certain trials. Of all the percent levels, 50% provides the most solid results, but logistically speaking a 30% sample is more attainable. However, even at the 10% level the systematic random sample still performs beautifully. At the beginning stage of an archaeological investigation, surface sampling is crucial to provide instruction as to which areas of the site may be significant. Due to this importance, systematic random sampling dominates its competitors as a way to provide archaeologists with a far-reaching distribution of samples, providing an overall better knowledge of the site. “This procedure has a number of advantages. Classes may be established with regard to different variables that one whish to control, which makes possible the reliable evaluation of variability in other phenomena with respect to the class-defined variables” (Binford 1964: 142).

At the 10% level, the phase percentage information is rather representative (see Figure 8). The 10% sample also provides a rather good physical distribution of units throughout the site (see map on following page with circled Tiwanaku ceramic clusters). Compared to the 10% random sample results, systematic random sampling provides a tremendously improved method. For the site of Pirque Alto, a 10% systematic sample returned results representative to the whole and would have been a useful method to employ.

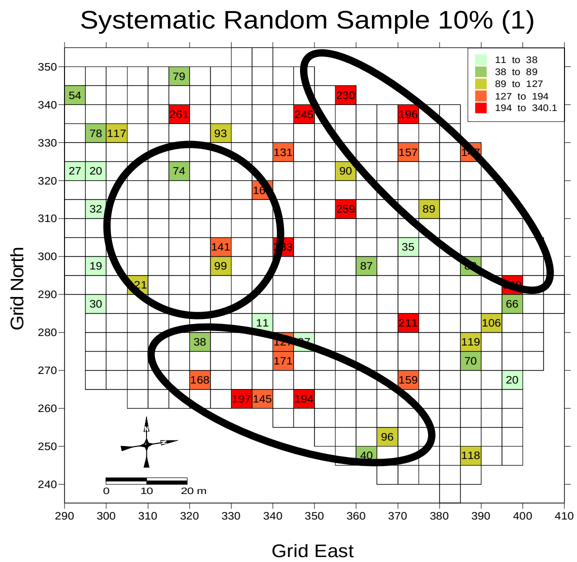

Figure 8. Systematic Random Sample 10% (Trial 1)

The 30% systematic random sample does an excellent job of displaying representative numerical results and gives excellent physical coverage. The circled areas on Figure 9 represent key Tiwanaku ceramic clusters.

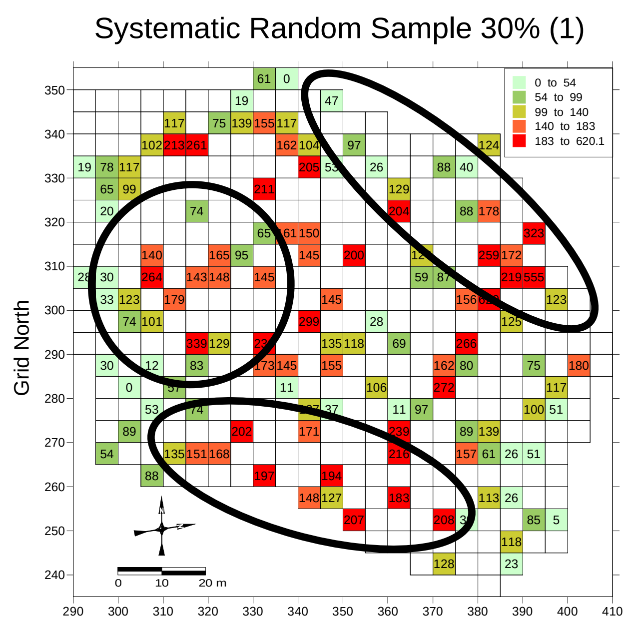

Figure 9. Systematic Random Sample 30% (Trial 1)

At the 50% level, the systematic random sample provides extremely close numeric results for all five trials. When spatially represented, this sample gives a very clear picture of areas of ceramic occupation on the site. In Figure 10, the Tiwanaku clusters have been circled.

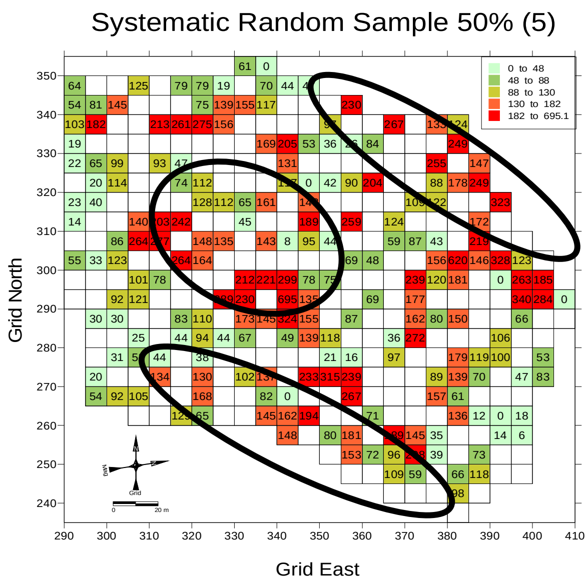

Figure 2. Systematic Random Sample 50% (Trial 5)

Table 2 shows the numeric results for all the systematic random trials shown above.

Table 2. Systematic Random Sample Numeric Results

	Whole Site	SRS 10% (Trial 1)	SRS 30% (Trial 1)	SRS 50% (Trial 5)
Number of Units Sampled	406	49	130	211
Mean ceramics per unit	128.81	119.92	128.56	124.25
Formative %	4.46%	4.61%	4.28%	3.96%
Early Intermediate %	0.90%	0.82%	0.94%	0.76%
Tiwanaku %	18.99%	19.84%	18.11%	19.00%
Diagnostic Ceramics %	25.24%	25.85%	24.24%	24.67%
Non-Diagnostic Ceramics %	74.76%	75.80%	75.76%	75.33%

Stratified Random Sampling. If random samples were the least spatially reliable in terms of finding identified clusters of ceramics, and systematic random samples were the best overall, stratified random sampling was firmly in the middle. Visually, the stratified samples performed better than the random samples (except at the 50% level, where they are roughly equal), but significantly less well than the systematic samples. You can see that, like the random sample at 10%, the stratified random sample tends to miss key clusters or cover them only very slightly. It is not until the 30% and especially 50% stratified sample level does this tend to correct itself slightly. Despite a fairly good coverage, there are still clusters of sampled units. However, Figure 11 for the most part offers a good spatial distribution.

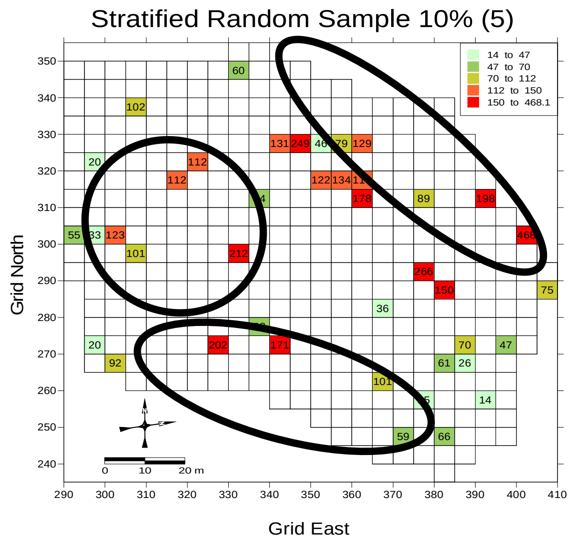

Figure 11. Stratified Random Sample 10% (Trial 5)

Figure 12 displays the visual distribution of the 30% sample level. The 30% sample level performed better in all trials than did the 10%.

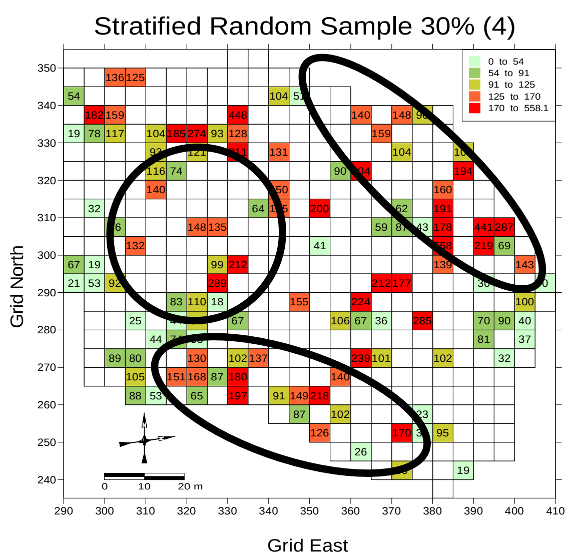

Figure 12. Stratified Random Sample 30% (Trial 4)

At the 50% level, (see Figure 13) the stratified sample, much like the systematic and the random become somewhat the same. While the high percentage guarantees a broad ranging recovery, it also starts to become redundant.

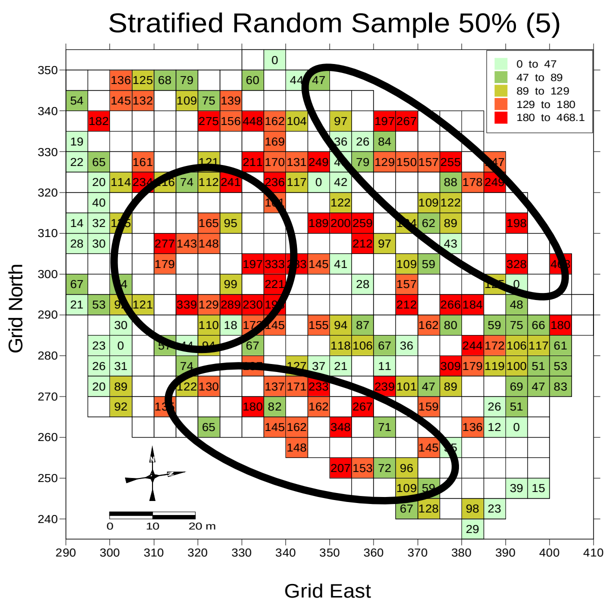

Figure 13. Stratified Random Sample 50% (5)

Table 3 displays the numeric results for all of the samples shown above.

Table 3. Stratified Random Sample Numeric Results

	Whole Site	Strat. 10% (Trial 5)	Strat. 30% (Trial 4)	Strat. 50% (Trail 5)
Number of Units Sampled	406	41	122	203
Mean ceramics per unit	128.81	109.24	121.28	121.39
Formative Phase %	4.46%	6.30%	4.85%	5.18%
Early Intermediate Phase %	0.90%	1.05%	0.95%	0.98%
Tiwanaku Phase %	18.99%	18.15%	21.85%	22.82%
Diagnostic Ceramics %	25.24%	26.35%	28.47%	29.91%
Non-Diagnostic Ceramics %	74.76%	73.65%	85.00%	70.09%

Judgmental Sampling. Numeric figures from all three judgmental samples (10, 30, and 50%) all provided representative results. (See table 4).

Table 4. Judgmental Sample Numeric Results

	Whole Site	Judgmental 10%	Judgmental 30%	Judgmental 50%
Number of Units Sampled	406	41	122	203
Mean ceramics per unit	128.81	130.63	148.63	131.04
Formative Phase %	4.46%	4.66%	4.66%	4.94%
Early Intermediate Phase %	0.90%	0.93%	0.93%	0.86%
Tiwanaku Phase %	18.99%	19.02%	19.02%	18.89%
Diagnostic Ceramics %	25.24%	25.54%	25.54%	25.59%
Non-Diagnostic Ceramics %	74.76%	74.46%	74.46%	74.41%

The judgmental samples had extremely varied spatial results. At the 10% level, the student chose a systematic sample approach and managed to cover the site rather uniformly. (See Figure 14).

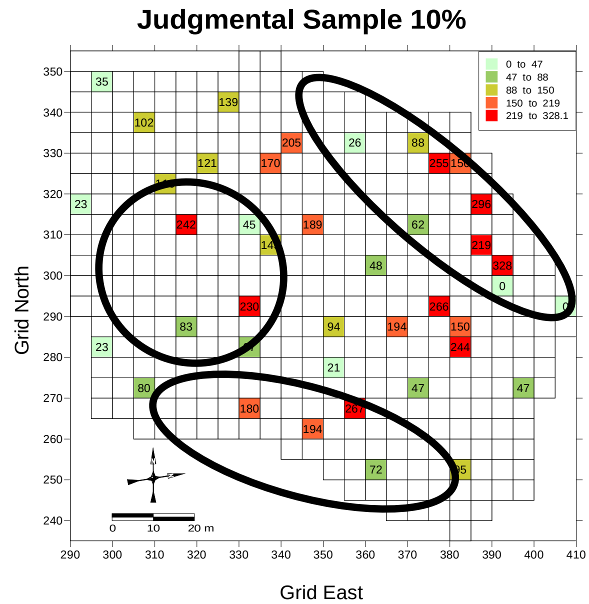

Figure 14. Judgmental Sample 10%

At the 30% level, the student chose a transect approach and cut the site into sections using three long corridors of sampled units. This student arranged the transects to intersect high, middle and low artifact areas, as well as to cover the visible features. (See Figure 15).

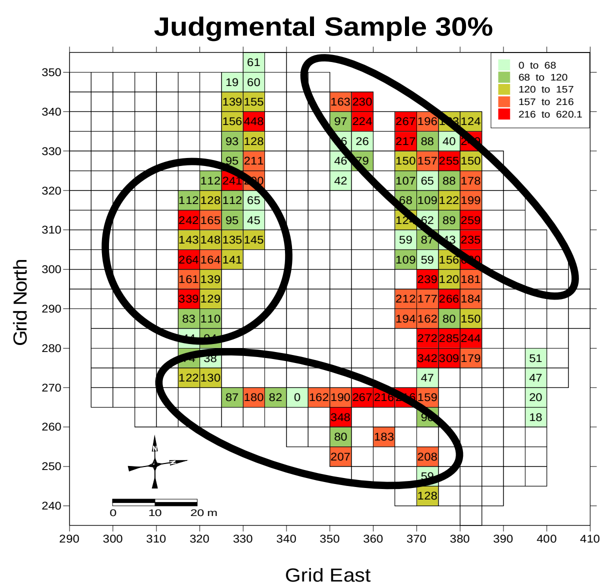

Figure 15. Judgmental Sample 30%

At the 50% level, the student again took a systematic approach; however, she avoided a property divider lined with cactus and thorny shrubbery. (See Figure 16).

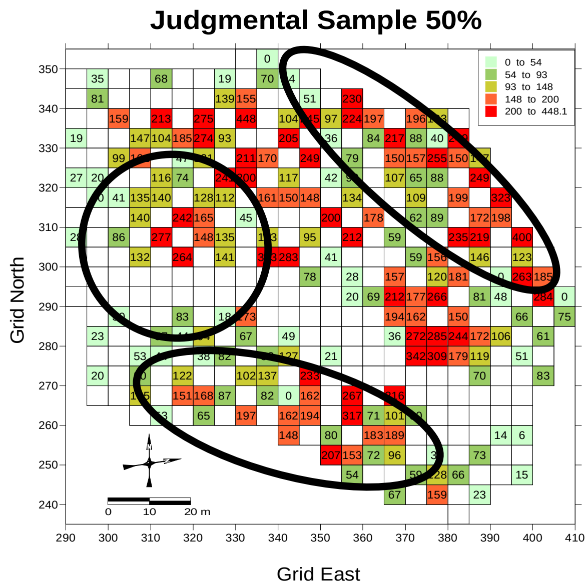

Figure 16. Judgmental Sample 50%

Visually, the 10% and 50% samples are quite similar. The 30% sample, featured in Figure 5 again with Tiwanaku clusters circled, presents a completely new method of sampling — transecting. Judgmental sampling has an extremely useful function in archaeology as it allows the field to become more flexible. Depending on the results of a systematic sample for example, judgmental sampling allows the archaeologist to redirect the project and investigate potential areas otherwise excluded through sampling.

Percent Interval

The most effective percent interval: 50%
The most realistic percent interval: 30%

Despite claims in the literature to the contrary, it is this author's belief that a 10% sample with any method is normally not conclusive or representative enough. What is being studied is comprised of many parts, some more measurable than others, and a 10% sample does not allow much leeway for the discovery of lesser reflected materials. However, should a 10% sample be the only available course of action, a systematic random sample should be the used, as at the 10% level its results were the most acceptable.

General Comments

Unfortunately, due to the nature and scale of this project, the role statistics and statistical estimations play in conjunction with sampling was not able to be explored. If someone were to continue the work begun by this project, the next logical step would be to conduct various statistical calculations (such as Chi square testing) to fully determine the statistical extent of the sample procedures.

CONCLUSIONS

At the start of this project, I found myself trying to straddle a line between the hardcore samplers with very complex math equations and those like Hole who saw sampling and statistics in archaeology as something that had focused too much energy on "probing the method instead of the ground" (Hole 1980: 219). After conducting my experiments and taking into account numeric and spatial representations and keeping in mind the various opinions within the literature, I still found myself straddling the line.

Through the process of conducting the experiments for this project and reading through the literature discussing the problems, concerns and benefits of sampling, it was possible to see how these factors played out within my experiments. At the end of this process, I had several more questions than answers and had grown immensely more cynical about math and sampling in regards to archaeology. I wrestled (on a much smaller scale) with all the issues mentioned in most of the journals and books I had read.

First, how does one go about choosing the appropriate sample size? “The optimum sample can be summarized as being small enough to yield statistically representative, significant, and accurate results at the minimum cost of labor” (Rootenberg 1964:186). For me, the choice was a relatively arbitrary one. I wanted to choose values that were both realistically applied archaeologically and that would give a varied enough result to draw conclusions about the various methods employed. However, when I examined the literature on determining appropriate sample size, there were many conflicting methods. Many were somewhat arbitrary, such as my own, i.e. believing or deciding that a 10% sample would be representative or a 20% sample, etc.

On the other end of the spectrum, I found some archaeologists recommended very complex mathematical or statistical equations in order to figure out the appropriate size. Looking at the math required, it seems likely that few archaeologists, without the help of a professional mathematician or statistician, will be able to do (and understand the implications of) the math required to determine sample size through these means. Approaching the topic as a realist, we need to work within our field to create methods that are logistically and practically applicable for the majority of individuals within our discipline.

The second issue I dealt with was if there is really any way to know that your sample is representative since archaeology deals with a seemingly indefinite population. Due to the nature of my thesis, I knew what my population was; I knew how many sherds total, I knew what phases, and I knew where. In the real world, no archaeologist has this advantage. The question that arose in my mind was this: is it worse to base your idea of what is representative on a complex mathematical equation, or draw upon similar archaeological sites combined with your knowledge of the area? Or is it better instead to use a combination, to rely upon your own judgment and knowledge of the area and then relate that to sampling and statistics as a way to aid you through the rest of the process, while still not being afraid to change your approach based on judgment?

The last of my four specific questions mentioned in the introduction dealt with the bias in sampling and if this was something able to be addressed or corrected. At the completion of this project, it occurred to me that the most damaging bias within sampling currently is the belief that with the right sampling method and the right sample size, a region or a site will be able to tell all. It is reflected in the literature that many put faith into these sampling methods that they no longer put into themselves. While Orton does allude to the field having returned to a more instinctive style of sampling, the effect that the explicit samplers has had on the field is irreversible (Orton 2000:6).

Is the way we conduct sampling intrinsically biased? To this I have to answer a resounding yes. Deciding the size of sample, the size of a unit, the boundaries of a site – all are typically made based upon the archaeologist’s judgment. Even if they are based with the most stringent of mathematics and statistics, you cannot simply cut and paste methods from one discipline into another. Statistics and sampling from any other field cannot be expected to operate in the same capacity within archaeology. “None of the sampling and estimation schemes currently in common use in archaeology take into account fully the spatial component in the data” (Hole 1980: 230). This is obviously something that needs to change. Bias is not necessarily a horrible thing, as long as we acknowledge it as a field, and adapt our methodology to take this into consideration.

For example, many site reports will simply state what percent sample was taken of the site, with very little regard to the process leading to that decision. If that discussion was included in the final site report, so that anyone who read it could understand why a 10%, or a 30% sample was chosen, or why a stratified approach was taken, then the harmful bias of sampling would be lessened. If you can legitimately explain your sampling strategy and how you came to that decision, the bias you have worked within becomes objective; consistency and specificity are crucial.

Based on the results of the sampling experiments conducted during this project and a careful review of the literature on the subject, the following conclusions have been reached: the most spatially and numerically representative sampling method is a systematic sample of at least 10% (though this author prefers a 30% sample). Numerically this method was not always as accurate as the random samples, however; it consistently at all percent levels provided a very even coverage of the site. As the results of surface recovery can be somewhat dubious (though it was decided at Pirque Alto the surface was a good indication of the subsurface) a systematic random sample approach on the subsurface would also prove extremely beneficial. At the beginning of an archaeological investigation, it would give you the best vantage point through which to view the site as a whole.

It is from this point that a judgmental sample is the next logical step. After the systematic random sample has provided wide-spread even coverage, the judgmental sample allows for the archaeologist to reevaluate the site, his or her research questions, and allows them the flexibility to decide where and how to proceed. This method should not be looked down upon or regarded as less effective or legitimate. The idea that all archaeology must be backed up by sophisticated mathematics and estimation statistics to be considered legitimate is rather insulting. “Without accurate assessment of the personal element, archaeology will be reduced to a set of mechanically applied approaches practiced by a series of competent but disinterested researchers, a grim prospect for the discipline” (Redman 1987: 262).

Sampling (and to some extent, statistics) within archaeology should be used as follows:

1. A method through which to locate archaeological sites on a regional scale.
2. A method through which to perform archaeology realistically, taking into account the current curation crisis, as well as monetary, labor and time constraints
3. An aid in the analysis of collected data, not as a means to do far too much with way too little. “Given the usual money and time restrictions, the minimum is far too often taken to be the maximum” (O’Neil 1993: 523).
4. A way to operate both spatially and numerically in a way to produce the most contextual results.

Though it is displeasing to end this thesis in a contradictory fashion, I find that it is where my experimental results and own personal conclusions lead. Sampling within archaeology is both at once a useful and dangerous tool. It allows us to work realistically within a world of constraints; however, it also has the propensity for great abuse. While my experiments concluded that a certain method of sampling at a certain percent was the most accurate and beneficial approach for the site of Pirque Alto, this does not mean that this is the approach that will work for every site; it would be naïve to think so.

Instead, the new direction of sampling in archaeology should be one of judgmental enforced flexibility. Sampling should continue to be relied upon, and the basic guidelines and the statistical analysis associated with it should be followed. However, it is extremely important that archaeologists expand on their methods, become far more specific about their decision making process, and not be afraid to rely on their own judgment. Archaeology is a complex, multifaceted field which requires more than an explicit methodology.

ACKNOWLEDGEMENTS

I would first like to thank Drs. Tim McAndrews and Connie Arzigian for their guidance throughout this project. Without them I would not have had a clue where to begin or how to proceed. I would also like to thank Dr. Jim Theler for redirecting my thoughts and questions. Thanks to the other members of the 2005 Bolivia Field School for helping to collect the raw data. Thanks to the Undergraduate Research Committee for funding this project. I would like to thank my mother and father Denise and Al Green for listening to me describe this project and giving me their support even though they are interested not in the slightest. I would like to acknowledge my friends, in particular Beth Plunger, Chris Salinas and Beth Boe for keeping me company throughout this process and offering their advice over much needed drinks. Lastly, I’d like to thank Phil Lee for his amazing proof reading abilities.

REFERENCES

Binford, Lewis. **1964**. A Consideration of Archaeological Research Design. *American Antiquity* 29:425-441.

Burkholder, JoEllen. **1997**. Tiwanaku and the Anatomy of Time: A New Ceramic Chronology From the Iwawi Site, Department of La Paz, Bolivia. New York: Binghamton University.

Drennan, Robert D. **1996**. *Statistics for Archaeologists: A Common Sense Approach.* New York: Plenum Press.

Flannery, Kent. **1970**. *The Early Mesoamerican Village*. New York: Academic Press.

Hole, Bonnie Laird. **1980**. Sampling in Archaeology: A Critique. *Annual Review of Anthropology* 9:217-234.

Janusek, John Wayne. **2004**. *Identity and Power in the Ancient Andes: Tiwanaku Cities Through Time*. New York: Routledge.

McAndrews, Timothy. **2006**. Preliminary Results from the Multicomponent Site of Pirque Alto in Cochabamba, Bolivia. Poster presented at the 71st annual meetings of the Society for American Archaeology in San Juan, Puerto Rico. Tim McAndrews, Claudia Rivera, Carla Jaimes.

Mueller, James. **1975**. *Sampling in Archaeology*. Tucson: The University of Arizona Press.

Nance, Jack. **1981**. Statistical Fact and Archaeological Faith: Two Models in Small-Sites Sampling. *Journal of Field Archaeology* 8:151-165.

O’Neil, Dennis. **1993**. Excavation Sample Size: A Cautionary Tale. *American Antiquity* 58:523-529.

Orton, Clive. **2000**. *Sampling in Archaeology*. Cambridge: Cambridge University Press.

Plog, Steven and Michelle Hegmon. **1993**. The Sample Size-Richness Relation: The Relevance of Research Questions, Sampling Strategies, and Behavioral Variation. *American Antiquity* 58:489-496.

Plog, Steven, Fred Plog, and Walter Wait. **1978**. Decision Making in Modern Surveys. In *Advances in Archeological Method and Theory*. Volume 1, edited by Michael Schiffer. pp. 383-421. Academic Press, Inc., New York.

Redman, Charles. **1987**. Surface Collection, Sampling and Research Design: A Retrospective. *American Antiquity* 52:249-265.

Renfrew, Colin, and Paul Bahn. **2000**. *Archaeology: Theories, Methods and Practice.* Third ed. New York: Thames & Hudson.

Rootenberg, S. **1964**. Archaeological Field Sampling. *American Antiquity* 30:181-188.

Schiffer, Michael, Alan P. Sullivan, Timothy Klinger. **1978**. The Design of Archaeological Surveys. *World Archaeology* 10:1-28.

Thompson, S.K. and G.A.F. Seber. **1996**. *Adaptive Sampling*. (New York: John Wiley and Sons).

Vescelius, G.S. **1960**. Archaeological Sampling: A Problem of Statistical Inference. In *Essays in the Science of Culture: In Honor of Leslie White*, edited by Gertrude E. Dole and Robert Carnerio. pp. 457-470. Crowell, New York.

Young – Sanchez, Maraget. **2004**. *Tiwanaku: Ancestors of the Inca*. University of Nebraska Press, Lincoln.

APPENDI

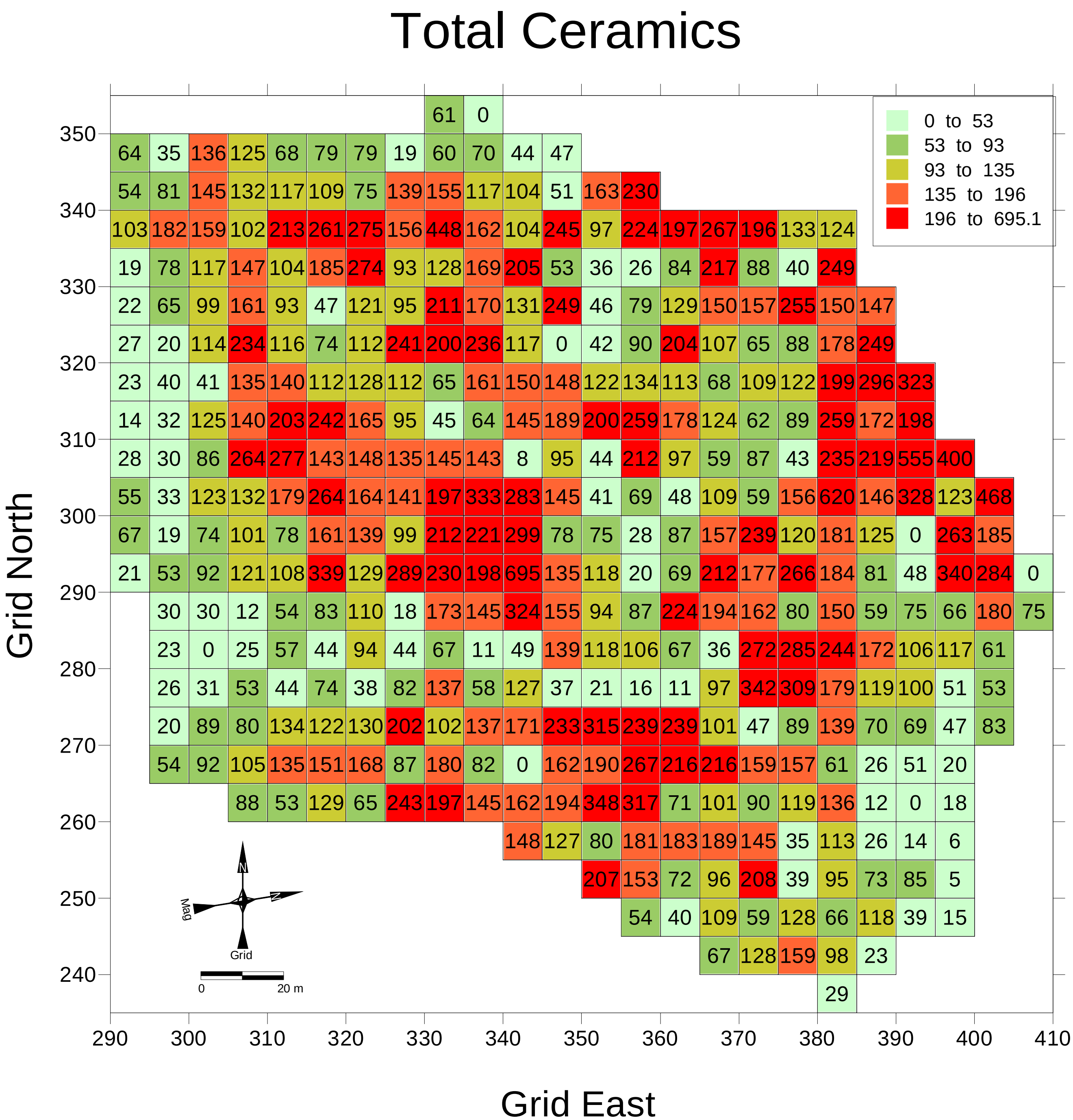

Figure 17 Total Ceramics Distribution

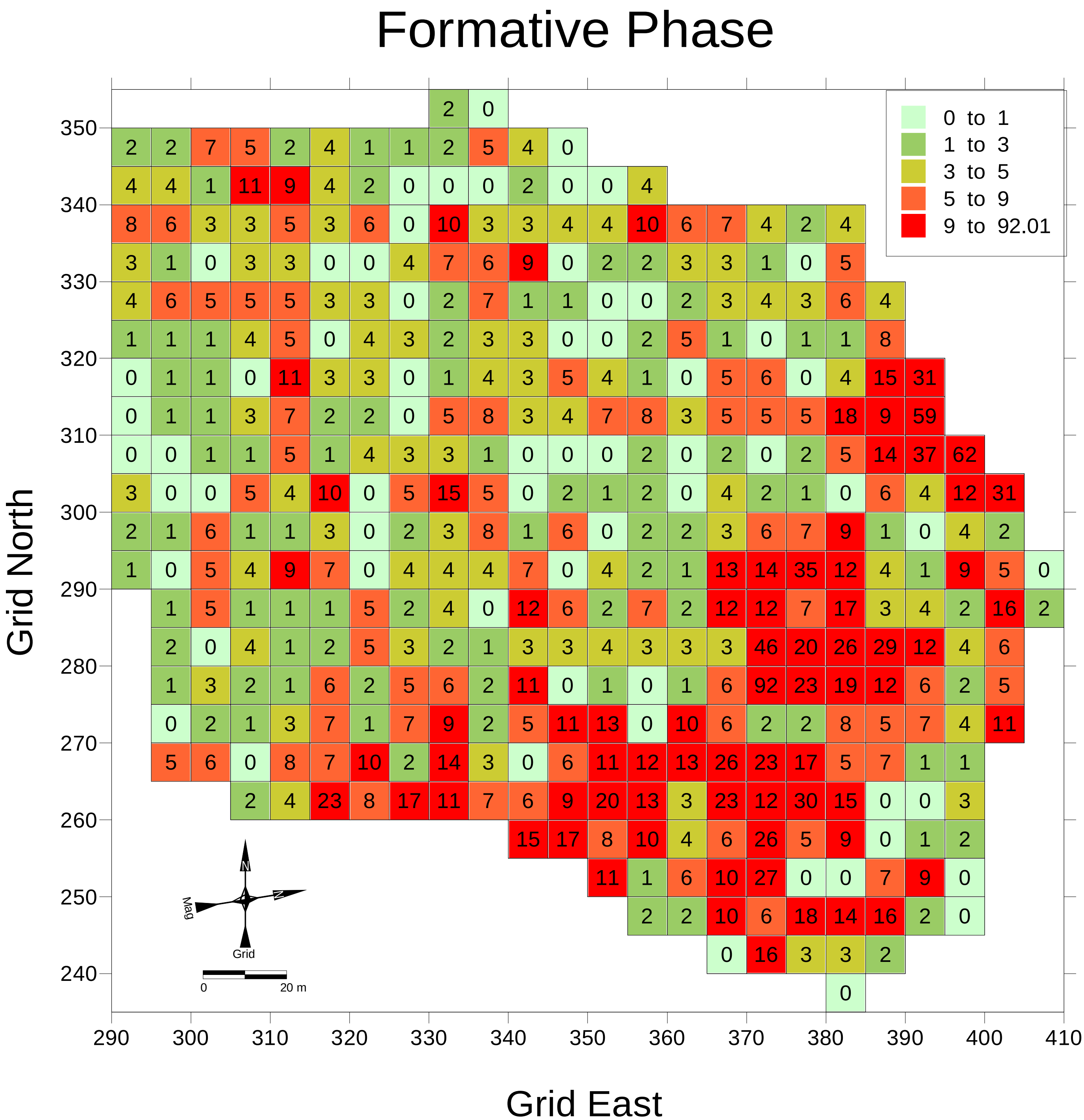

Figure 18 Formative Phase Ceramic Distribution

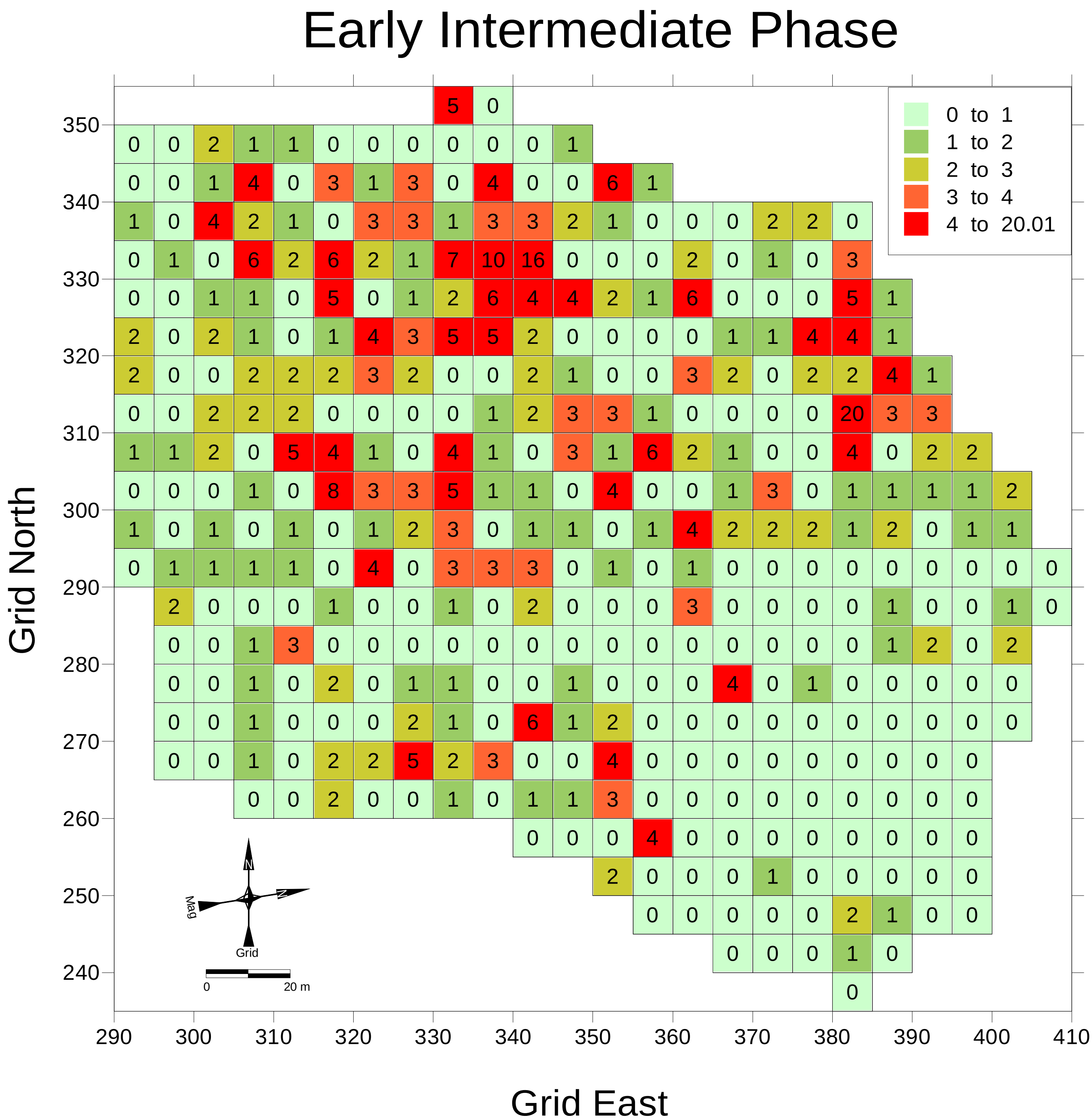

Figure 19 Early Intermediate Phase Ceramic Distribution

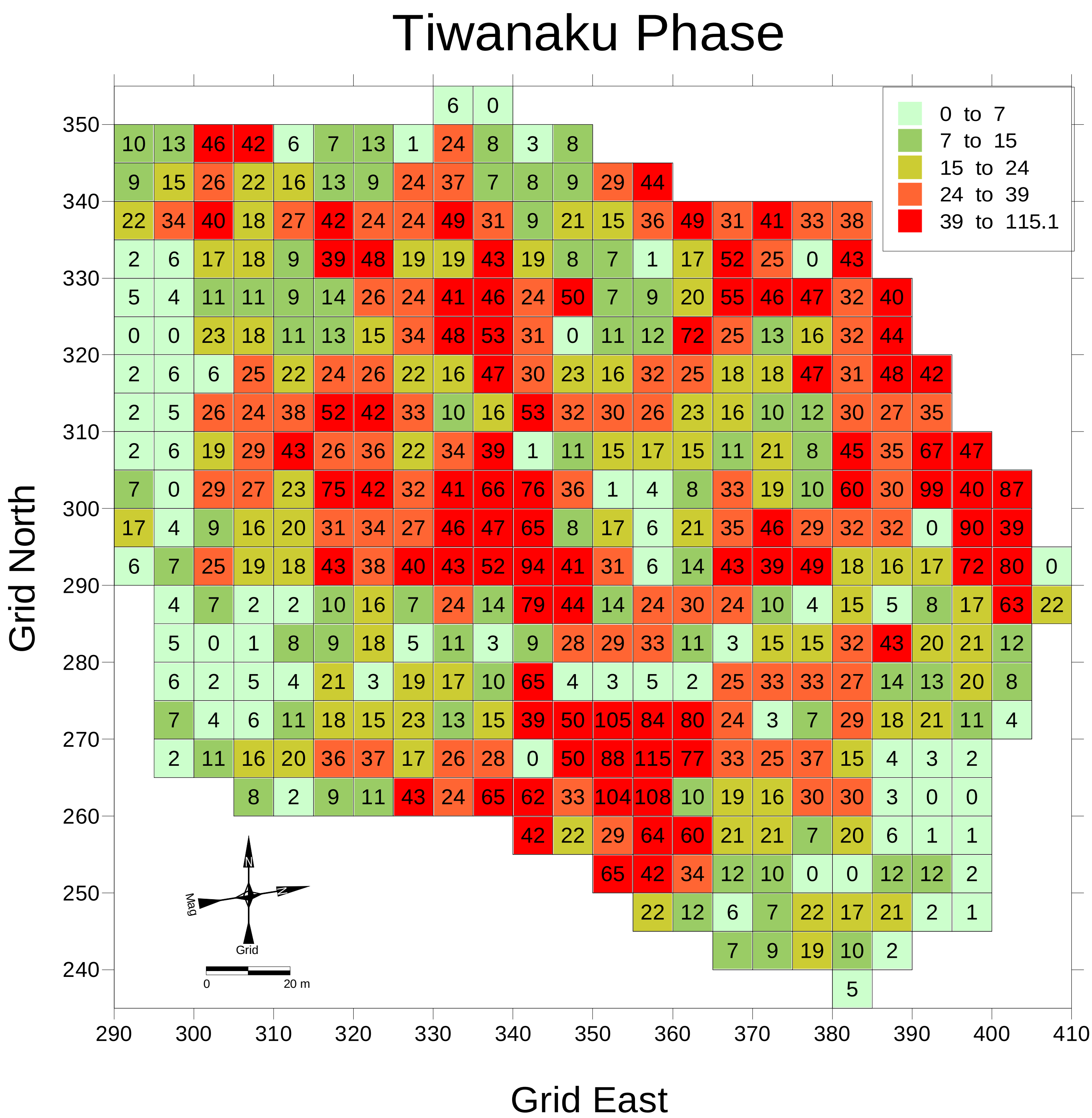

Figure 20 Tiwanaku Phase Ceramic Distribution