ELSEVIER

Journal of Archaeological Science 34 (2007) 1985–1990

Journal of Archaeological SCIENCE

http://www.elsevier.com/locate/jas

On the use of species-area curves to detect the effects of sample size

R. Lee Lyman [a,*], Kenneth M. Ames [b]

[a] *Department of Anthropology, University of Missouri, Columbia, MO 65211, USA*
[b] *Department of Anthropology, Portland State University, Portland, OR, USA*

Received 7 December 2006; received in revised form 23 January 2007; accepted 24 January 2007

Abstract

Archaeologists have long been concerned about sample size. To monitor the relationship between sample size and the magnitude of a variable of interest, archaeologists have constructed three distinct kinds of what in ecology are known as species area curves (SACs). Sampling to redundancy SACs involve adding successive samples cumulatively to determine if the information provided by new samples is unique or redundant with information provided by earlier (smaller) samples. Rarefaction SACs involve pooling multiple samples and then rarifying the pooled sample—making it smaller probabilistically—so that the originally large but rarified sample may be compared with originally small samples. Regression SACs determine if there is a significant statistical relationship between samples of different sizes and the magnitude of the variable of interest per sample. Each of the three SACs used by archaeologists has a distinct analytical purpose.

Keywords: Rarefaction; Regression; Sample-size effect; Sample to redundancy; Species area curve

1. Introduction

Ever since the 1960s when archaeologists recognized that their samples of artifacts and ecofacts might not be representative of a target variable of interest, they have worried about how to determine if, or ensure that, they had representative samples (e.g., Mueller, 1975). Some years after the concern for sample adequacy had been identified, it was pointed out that the values of many of the quantitative variables archaeologists measure were potentially the result of sample size or the frequency of phenomena tallied (Grayson, 1981; Jones et al., 1983; Thomas, 1983). This in turn prompted a plethora of studies that explored the relationship between sample size and a quantitative variable of interest, most often the number or richness of kinds represented in a sample (e.g., Grayson, 1989, 1991; Jones et al., 1989; Kirch et al., 1987; Leonard, 1989a,b; Lyman, 1991; Meltzer et al., 1992; Shott, 1989; Simek and Snyder, 1988). Ultimately, methods were designed to estimate richness given a particular sample size (e.g., Kintigh, 1984); these methods were critiqued (Rhode, 1988) and modified (McCartney and Glass, 1990), and new methods were proposed (e.g., Kaufman, 1998).

Various methods of studying the relationship between samples of different sizes and a variable of interest continue to be used (e.g., Byrd, 1997), evaluated (e.g., Ringrose, 1993; Sullivan and Tolonen, 1998), modified (e.g., Baxter, 2001), and developed (e.g., Cannon, 2001; Zohar and Belmaker, 2005). A half-dozen introductory archaeology textbooks published in the last five years all discuss probabilistic sampling, but none mentions any of the ways that sample adequacy might be and indeed has been evaluated by archaeologists over the past two and a half decades. Nor do any of those textbooks discuss how archaeologists have studied the relationship between sample size and a variable of interest. This means that numerous case studies must be examined if a novice wishes to gain a basic

* Corresponding author.
E-mail address: lymanr@missouri.edu (R.L. Lyman).

0305-4403/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jas.2007.01.011

understanding of the methods used. We hope to here reduce the magnitude of that requirement.

We recently explored the sampling to redundancy (STR, hereafter) protocol using archaeofaunal remains (Lyman and Ames, 2004). In doing so, we realized, but did not comment on the fact that what is known generically in the ecological literature as a species-area curve actually comprises several distinct kinds of relationship (Gotelli and Colwell, 2001). Our purposes here are to illustrate three distinct kinds of species-area curve (SAC, hereafter) that have been used by archaeologists and to argue that each kind serves a unique analytical purpose. We do this for two reasons. First, we are unaware that the distinctions we make here have been previously made explicit in one place in the archaeological literature. Rather, in that literature only one kind of curve typically is used in any given study, hence the different kinds of curves may be conceived as equivalent across several different articles and may result in the impression that the different kinds of SACs measure the same relationship (e.g., Lepofsky and Lertzman, 2005). It seems to us that such potential confusion is easily avoided if differences are identified and explicated. And second, we believe that there must be a link between analytical problem or research question and the method used to solve the problem or answer the question. By clarifying these distinctions, we hope to help archaeologists select the analytically appropriate curve.

2. Background to species-area curves

Plant ecologists recognized more than 100 years ago that as the area sampled increased in size, so too did the number of species found (e.g., Arrhenius, 1921; Evans et al., 1955; Fisher et al., 1943; Gleason, 1922). Today the relationship is, according to ecologist T.W. Schoener (1986: 560), ''one of community ecology's few universal regularities.'' Not surprisingly, then, ecologists continue to grapple with building methods to compare species richness values across communities (Gotelli and Colwell, 2001). Perhaps what is surprising, at least to an archaeologist or paleontologist (e.g., Barnosky et al., 2005), is the recent recognition by ecologists that not only does an increase in the sampled amount of geographic space result in more species, but so too does the amount of time sampled (Adler and Lauenroth, 2003).

The construction of a SAC involves plotting a measure of sampling effort—whether area surveyed, number of samples examined, individuals tallied, hours spent searching, or the like—on the *x*-axis of a bivariate scatterplot, and a measure of the ecological property of interest—taxonomic richness, evenness, heterogeneity—on the *y*-axis. Given that populations of many phenomena are finite and tend to have skewed frequency distributions such that some kinds are abundant, other kinds are not so abundant, and still other kinds are quite rare, the SAC will often be initially steep but gradually level off as common taxa are found first and successively more rare taxa are found only with successively larger samples. That is, theory and empirical data indicate that as sample size increases, new or unique taxa are initially added quickly (because they are abundant) and are added progressively more slowly (because they are rare) as sample size increases (Jones et al., 1983).

As implied in the preceding paragraph, sampling effort can be measured in one of two basic ways: as the number of individuals, or as the number of samples. A sample may consist of a unit of space, a unit of time, or some other entity in which ≥ 0 individuals can occur. SACs may be constructed in one of two basic ways: as accumulations of individuals or samples, or as rarefaction curves. The former involve adding successively collected individuals or samples to form a single large collection. The latter generally involve probabilistically sampling a large pool of individuals or samples, at random, generally without replacement (Gotelli and Colwell, 2001; see Scheiner, 2003 for additional variations). Archaeologists have constructed SACs using individuals and using samples, and they have used versions of both curve-building procedures, as well as a less dynamic procedure to build SACs.

In the remainder of this discussion, we assume that the target variable is taxonomic richness, or the number of taxa (NTAXA) in a collection. Whether those taxa are animal species, plant species, or types of artifacts is irrelevant. As in community ecology, the greater the size of the sample (whether number of individuals or number of samples) of artifacts or other archaeological phenomena, the greater the value of NTAXA, usually.

3. Kinds of species-area curves

The SACs that archaeologists, zooarchaeologists, and ethnobotanists (and ecologists in general) have created have been built in one of three ways. One way to construct a SAC is to sample to redundancy; add successive samples together such that the *x*-axis measures a cumulatively pooled series of samples, resulting in statistically interdependent samples. Archaeologists refer to this analytical protocol as the ''sample to redundancy'' (STR hereafter) procedure (Dunnell, 1984; Leonard, 1987). As new samples or individuals are added successively to a collection, the value of the target variable is monitored. When the value of the target variable is stable across several successive sample additions, so the reasoning goes, one can argue that the total sample is representative of the target variable because all information in new samples is redundant with information provided by the previous (smaller) total sample. If the STR curve does not stabilize and level off (the target variable continues to change value with each new sample), then the samples are not redundant, the total combined sample is not representative of the target variable, and all taxonomic variation in the target population has not yet been sampled.

When the STR protocol is used, a curve can be visually fit to data points (e.g., Wolff, 1975), or it can be fit statistically (e.g., Lepofsky et al., 1996). An example of a visually fit STR curve is shown in Fig. 1 and is based on the mammalian genera data for the Meier site (35CO5) in northwestern Oregon State that is summarized in Table 1 (Lyman and Ames, 2004). In the case of Fig. 1 the STR curve levels off, so we

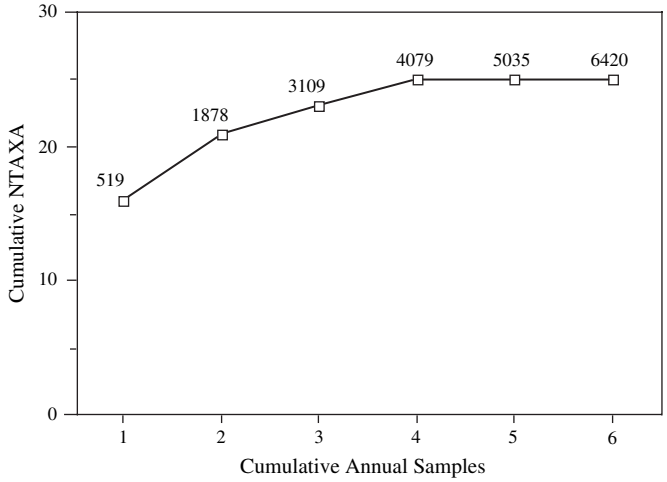

Fig. 1. Cumulative richness of mammalian genera across cumulative incremental samples from the Meier site. Cumulative sample size (∑NISP) is indicated per sample. Data from Table 1.

conclude that we have sampled to redundancy (at least with respect to non-rare taxa) because all taxa encountered in new samples have been encountered in earlier (smaller total) samples (Lyman and Ames, 2004). The STR protocol is a dynamic analytical procedure—add or accumulate individuals or samples—that has as its purposes allowing one to determine if, and to empirically demonstrate that, the total collection is representative of the variable of interest. We have found few archaeological applications of the STR protocol in the archaeological literature (e.g., Lepofsky et al., 1996; Lepofsky and Lyons, 2003; Lyman, 1991), though we did find what seems to be an independent invention of the technique (Monks, 2000).

It is important to note that the order in which samples are added will likely influence the sample size at which the STR curve levels off (Kerrich and Clarke, 1967). We have added annual samples in the chronological order in which they were collected (e.g., Fig. 1) in an effort to determine if additional years of excavation were required (Lyman and Ames, 2004). We have also added samples in order from largest to smallest on the presumption that larger samples would be richer than smaller samples and thus the STR curve would level off more quickly as the accumulated added samples became progressively smaller (Lyman and Ames, 2004). In short, an explicit justification for the order in which samples are added in the STR procedure is recommended.

The second way that SACs have been built by archaeologists involves a procedure termed *rarefaction* by ecologists (e.g., Sanders, 1968; Tipper, 1979). It involves reducing sample sizes probabilistically in order to compare samples of quite different sizes by making their sizes similar (e.g., Kintigh, 1984; McCartney and Glass, 1990; Shaw, 1995). Knowledge of tallies per taxon are required for rarefaction but not for STR or the regression approach (see below). Rarefaction has been known in zooarchaeology (Styles, 1978, 1981) and paleoethnobotany (Fasham and Monk, 1978) for nearly 30 years, and in archaeology for more than two decades (e.g., Kintigh, 1984). It has been used recently by zooarchaeologists (e.g., Byrd, 1997) and by ethnobotanists (e.g., Lepofsky and Lertzman, 2005), despite earlier arguments that the probability of specimen interdependence invalidates application of the procedure to these materials (Grayson, 1984: 152). Counting individual bone and tooth or seed and twig specimens may result in a single organism (animal or plant, respectively) being counted more than once. The same argument would likely not apply to stone artifacts (assuming they were complete and not fragmentary) though it likely would apply to ceramic sherds if the variable of interest were NTAXA.

There are several ways to construct rarefaction curves, but describing them is beyond our scope here. It suffices to say that one can use statistically independent samples or statistically interdependent (summed) samples (or sample without replacement, or sample with replacement) to estimate taxonomic richness where a sample is of a particular size. An example of a SAC of the rarefaction kind, using the six annual samples from the Meier site described in Table 1, is presented in Fig. 2. To generate this curve, we used Steve Holland's (2005) software for "Analytical Rarefaction" which is based on Raup (1975) and Tipper (1979). The rarified sample and richness values in Fig. 2 allow us to compare the modeled richness of the Meier assemblage with the richness of another assemblage that is considerably smaller than the Meier assemblage without fear of differences in sample size obscuring results or producing false results (if we ignore problems of specimen interdependence). And that

Table 1
NISP (number of identified specimens) of mammalian genera for each year of excavation at the Meier site (35CO5)

Genus	1973	1987	1988	1989	1990	1991	Total
Scapanus	4	4	3	4	1	2	18
Sylvilagus	2	3	1	1	10	1	18
Aplodontia	2	1	0	1	0	3	7
Eutamias	0	0	0	1	0	0	1
Tamiasciurus	0	0	2	0	0	0	2
Thomomys	0	2	0	1	5	1	9
Castor	13	100	65	52	41	71	342
Peromyscus	0	4	12	12	4	3	35
Neotoma	0	0	1	0	0	0	1
Microtus	0	15	25	34	15	11	100
Ondatra	37	97	55	59	74	52	374
Erethizon	0	0	0	1	0	0	1
Canis	21	25	13	16	11	25	111
Vulpes	3	1	1	0	0	0	5
Ursus	20	16	20	7	13	26	102
Procyon	15	79	51	35	43	64	287
Martes	1	6	1	0	1	11	20
Mustela	4	35	17	19	38	21	134
Mephitis	0	1	1	0	0	2	4
Lutra	6	12	6	2	11	14	51
Felis	0	4	1	0	3	1	9
Lynx	9	5	4	1	4	8	31
Phoca	3	6	5	10	6	13	43
Cervus	103	165	191	152	106	218	935
Odocoileus	276	788	756	562	570	838	3780
Total NISP	519	1359	1231	970	956	1385	6420
NTAXA	16	21	21	19	18	20	25
Cumulative NISP	519	1878	3109	4079	5035	6420	
Cumulative NTAXA	16	21	23	25	25	25	

From Lyman and Ames (2004: 336).

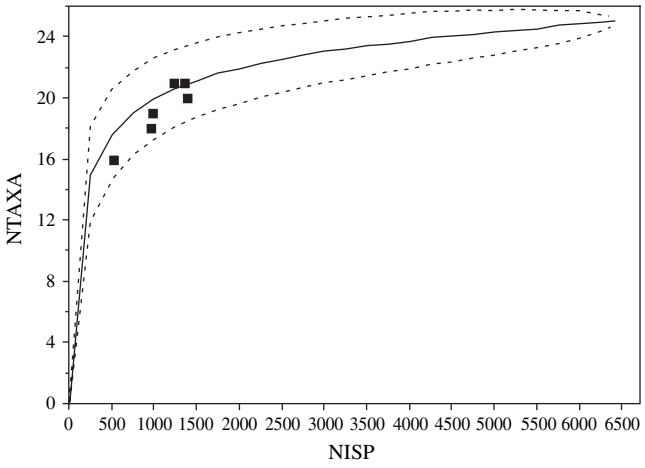

Fig. 2. Rarefaction curve (solid line) and 95 percent confidence intervals (dashed lines) of richness of mammalian genera based on six annual samples from the Meier site (black squares). Data from Table 1; curves estimated using Analytical Rarefaction in Holland (2005).

is the purpose of SACs constructed as rarefaction curves (e.g., Kintigh, 1984), a purpose distinct from that attending SACs built using the STR procedure.

The third way that SACs are constructed is the least dynamic, in our view. It involves generating bivariate plots of statistically independent samples of different sizes against their respective richness values, and then statistically fitting a curve to the plot in order to determine if sample size is influencing NTAXA. An example is shown in Fig. 3, using the same six annual samples from the Meier site described in Table 1. The best-fit regression line is statistically significant ($p < .01$) and suggests that NTAXA per statistically independent annual sample is a function of sample size measured as NISP. If each point represented a sample from a different stratum or a different site, Fig. 3 would suggest that those samples were strongly influenced by sample size and thus NTAXA values for those samples should not be compared.

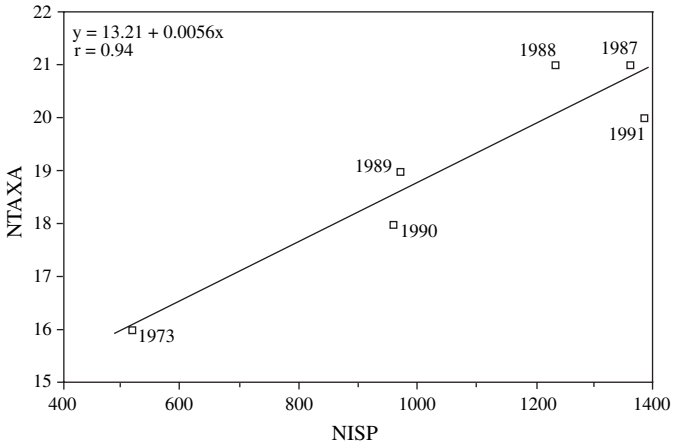

Fig. 3. Bivariate scatterplot of sample size (NISP) and richness of mammalian genera across six annual samples from the Meier site. The simple best-fit regression line is shown; formula describing that line is given ($p < .01$). Year when a sample was collected is indicated. Data from Table 1.

The technique of building a SAC exemplified in Fig. 3 is often referred to as the "regression method" (e.g., Baxter, 2001; Leonard, 1997). It is this method of building an SAC that caused archaeologists to realize that what became known as the "sample size effect" could be a problem (Thomas, 1983; for examples, see Bobrowsky, 1982; Grayson, 1978, 1984; Hesse, 1982; Jones et al., 1983; Kirch et al., 1987; Leonard, 1989a,b; Lyman, 1991; Sharp, 1990). The analytical purpose of constructing such a curve is to determine if the values of a target variable manifest by samples of different sizes might be compared without fear of sample size effects. If there is a strong correlation between sample size and the property of interest, the typical conclusion is that the magnitudes of the variable of interest could be a function of sample size. Although it may not be the case that the values of the target variable vary purely as a result of sample size (e.g., Baxter, 2001; Dunnell, 1989; Plog and Hegmon, 1993, 1997), it is a potentiality well worth considering (Leonard, 1997), and the purpose of the regression approach is to detect when such sample-size effects may be present.

If sample-effects are suggested by regression analysis, the analyst has options. The samples may be pooled and a rarefaction analysis performed (e.g., Kintigh, 1984). Alternatively, slopes of lines describing the relationship between sample size and NTAXA may be found to vary between different sets of samples (e.g., Grayson, 1991, 1998; Lepofsky et al., 1996). If so, comparisons of the slopes may reveal a property of the sets of collections not otherwise detectable and, more importantly, avoids sample-size effects. Another use of the regression approach has been to identify outliers, or assemblages that fall significant distances (e.g., >2 standard deviations) from the regression line (e.g., Grayson, 1984). Determination of why such outlying assemblages do not fit the general sample-size driven pattern may reveal unique properties of the unusual assemblage; it may, for example, be derived from a different population than the other samples. Study of multiple slopes of best-fit lines and of outliers revealed by the regression approach both avoid some weaknesses of the approach such as the fact that some relatively small samples may in fact be 100 percent samples (e.g., Baxter, 2001; Rhode, 1988; Sullivan and Tolonen, 1998). Regardless, the regression approach of building SACs obviously has a different analytical purpose(s) than the curves based on STR and curves based on rarefaction.

4. Discussion

The three curves shown in Figs. 1–3 vary in general appearance despite the fact that the data used to build them all derive from Table 1. They differ in appearance because they are meant to address different questions and thus are built differently. The STR curve (Fig. 1) allows determination of one total collection's (a set of pooled samples) representativeness and can suggest when sufficient excavation, collection, and analysis have been accomplished (Lyman and Ames, 2004). Rarefaction (Fig. 2) allows two or more samples of different sizes to be compared by analytically making them the same size. Regression analysis (Fig. 3) detects possible

sample-size effects among samples of different sizes, and can lead to additional analyses. The three types of SACs reveal different properties of the relation between a set of samples (of similar or quite diverse sizes) and a target variable.

The regression approach may be applied to samples that share taxa or samples that do not share taxa, as may rarefaction. However, it seems to us that all three approaches to building SACs discussed here are most usefully applied to samples thought to have been derived from the same population and thus share some taxa. If samples are not from the same population and therefore share few or no taxa, then questions of sample adequacy that originally drove (and continue to drive such analyses) become murky if they are indeed still applicable.

Use of individual samples of varying size rather than equal tallies of individuals per sample (individual-based samples) on the *x*-axis produces a measurement of "species density" (NTAXA per NISP in a sample) instead of a measurement of the influence of sample size (tallies of individuals) on NTAXA (Lepofsky and Lertzman, 2005: 183; see also Colwell et al., 2004; Gotelli and Colwell, 2001). Because species density may be context specific, use of samples comprising ≥ 0 individuals rather than a standard number of individuals can produce "erroneous conclusions" regarding the relationship between sample size and NTAXA (Lepofsky and Lertzman, 2005: 183). We suggest, however, that if one is interested in whether or not a total collection is representative of NTAXA, one can use individual samples of varying size to construct an STR curve without fear of measuring species density (Lyman and Ames, 2004). This is so because the analyst is not comparing NTAXA across samples of different sizes, and so is not comparing species densities. Rather, one is using an STR curve for its analytical purpose of assessing representativeness of a collection comprising multiple samples (Jamniczky et al., 2003; Wolff, 1975).

5. Conclusion

There are many valid uses of the several kinds of SACs, but the particular kind of SAC used should attend a particular analytical purpose. The studies cited here, and many others not cited, have concerned the species-area relationship in one of the forms we have described, but with few exceptions, they have not considered the differences between the kinds of SACs that we have described. No archaeologist we are aware of has described in detail the particular analytical purposes of the three different kinds of SACs discussed here. Also, with few exceptions, only one kind of species-area curve has been built in any given publication. One notable exception is found in paleontology where Jamniczky et al. (2003) built STR curves, and then superimposed rarefaction curves they had constructed using pooled data over the STR curves to assess goodness of fit of the two and to assist with evaluating the validity of the STR curves. Another exception is Lepofsky et al.'s (1996) use of both STR and regression curves to compare the floral assemblages in three prehistoric houses.

By identifying the different kinds of SACs, and highlighting their differences by constructing different SACs using the same data set, we hope to have made clear that each kind of SAC focuses on a specific analytical problem. The history of archaeology is full of examples of vague use of poorly defined labels resulting in confusion and misunderstanding (e.g., Lyman et al., 1998). There should no longer be a reason to confuse, or mislabel, the different kinds of SACs.

Acknowledgments

We thank Virginia Butler, Todd VanPool, and several anonymous reviewers for comments on early drafts.

References

Adler, P.B., Lauenroth, W.K., 2003. The power of time: spatiotemporal scaling of species diversity. Ecology Letters 6, 749–756.

Arrhenius, O., 1921. Species and area. Journal of Ecology 4, 68–73.

Barnosky, A.D., Carrasco, M.A., Davis, E.B., 2005. The impact of the species-area relationship on estimates of paleodiversity. PLoS Biology 3, 1356–1361.

Baxter, M.J., 2001. Methodological issues in the study of assemblage diversity. American Antiquity 66, 715–725.

Bobrowsky, P.T., 1982. An examination of Casteel's MNI behavior analysis: a reductionist approach. Midcontinental Journal of Archaeology 7, 173–184.

Byrd, J.E., 1997. Quantifying hunter-gatherer intensification: a zooarchaeological case from Arctic Canada. Journal of Anthropological Archaeology 16, 49–72.

Cannon, M.D., 2001. Archaeofaunal relative abundance, sample size, and statistical methods. Journal of Archaeological Science 28, 185–195.

Colwell, R.K., Mao, C.X., Chang, J., 2004. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. Ecology 85, 2717–2727.

Dunnell, R.C., 1984. The ethics of archaeological significance decisions. In: Green, E.L. (Ed.), Ethics and Values in Archaeology. Free Press, New York, pp. 62–74.

Dunnell, R.C., 1989. Diversity in archaeology: a group of measures in search of an application? In: Leonard, R.D., Jones, G.T. (Eds.), Quantifying Diversity in Archaeology. Cambridge University Press, Cambridge, pp. 142–149.

Evans, F.C., Clark, P.J., Brand, R.H., 1955. Estimation of the number of species present on a given area. Ecology 36, 342–343.

Fasham, P.J., Monk, M.A., 1978. Sampling for plant remains from Iron Age pits: some results and implications. In: Cherry, J.F., Gamble, C., Shennan, S. (Eds.), Sampling in Contemporary British Archaeology. BAR British Series 50, Oxford, pp. 363–371.

Fisher, R.A., Corbet, A.S., Williams, C.B., 1943. The relation between number of species and the number of individuals in a random sample of animal population. Journal of Animal Ecology 12, 42–58.

Gleason, H.A., 1922. On the relation between species and area. Ecology 3, 158–162.

Gotelli, N.J., Colwell, R.K., 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology Letters 4, 379–391.

Grayson, D.K., 1978. Minimum numbers and sample size in vertebrate faunal analysis. American Antiquity 43, 53–65.

Grayson, D.K., 1981. The effect of sample size on some derived measures in vertebrate faunal analysis. Journal of Archaeological Science 8, 77–88.

Grayson, D.K., 1984. Quantitative Zooarchaeology. Academic Press, Orlando.

Grayson, D.K., 1989. Sample size and relative abundance in archaeological analysis: illustrations from spiral fractures and seriation. In: Leonard, R.D., Jones, G.T. (Eds.), Quantifying Diversity in Archaeology. Cambridge University Press, Cambridge, pp. 79–84.

Grayson, D.K., 1991. Alpine faunas from the White Mountains, California: adaptive change in the late prehistoric Great Basin? Journal of Archaeological Science 18, 483–506.

Grayson, D.K., 1998. Moisture history and small mammal community richness during the latest Pleistocene and Holocene, northern Bonneville Basin, Utah. Quaternary Research 49, 330–334.

Hesse, B., 1982. Bias in the zooarchaeological record: suggestions for interpretation of bone counts in faunal samples from the Plains. In: Ubelacker, D.H., Viola, H.J. (Eds.), Plains Indian Studies: A Collection of Essays in Honor of John C. Ewers and Waldo R. Wedel. Smithsonian Contributions to Anthropology No. 30, Washington, D.C., pp. 157–172.

Holland, S. 2005. Analytical rarefaction. Version 1.3. http://www.uga.edu/~strata/software/software.html, verified Nov. 15, 2006.

Jamniczky, H.A., Brinkman, D.B., Russell, A.P., 2003. Vertebrate microsite sampling: how much is enough? Journal of Vertebrate Paleontology 23, 725–734.

Jones, G.T., Grayson, D.K., Beck, C., 1983. Artifact class richness and sample size in archaeological surface assemblages. In: Dunnell, R.C., Grayson, D.K. (Eds.), Lulu Linear Punctated: Essays in Honor of George Irving Quimby. Anthropological Papers No. 72. Museum of Anthropology, University of Michigan, Ann Arbor, pp. 55–73.

Jones, G.T., Beck, C., Grayson, D.K., 1989. Measures of diversity and expedient lithic technologies. In: Leonard, R.D., Jones, G.T. (Eds.), Quantifying Diversity in Archaeology. Cambridge University Press, Cambridge, pp. 69–78.

Kaufman, D., 1998. Measuring archaeological diversity: an application of the jackknife technique. American Antiquity 63, 73–85.

Kerrich, J.E., Clarke, D.L., 1967. Notes on the possible misuse and errors of cumulative percentage frequency graphs for the comparison of prehistoric artefact assemblages. Proceedings of the Prehistoric Society 33, 57–69.

Kintigh, K.W., 1984. Measuring archaeological diversity by comparison with simulated assemblages. American Antiquity 49, 44–54.

Kirch, P.V., Allen, M.S., Butler, V.L., Hunt, T.L., 1987. Is there an early Far Western Lapita province? Sample size effects and new evidence from Eloaua Island. Archaeology in Oceania 22, 123–127.

Leonard, R.D., 1987. Incremental sampling in artifact analysis. Journal of Field Archaeology 14, 498–500.

Leonard, R.D., 1989a. Anasazi Faunal Exploitation: Prehistoric Subsistence on Northern Black Mesa, Arizona. Center for Archaeological Investigations Occasional Paper No. 13. Southern Illinois University Press, Carbondale.

Leonard, R.D., 1989b. Resource specialization, population growth, and agricultural production in the American Southwest. American Antiquity 54, 491–503.

Leonard, R.D., 1997. The sample size-richness relation: a comment on Plog and Hegmon. American Antiquity 62, 713–716.

Lepofsky, D., Lertzman, K., 2005. More on sampling for richness and diversity in archaeobiological assemblages. Journal of Ethnobiology 25, 175–188.

Lepofsky, D., Lyons, N., 2003. Modeling ancient plant use on the Northwest Coast: towards an understanding of mobility and sedentism. Journal of Archaeological Science 30, 1357–1371.

Lepofsky, D., Kusmer, K., Hayden, B., Lertzman, K., 1996. Reconstructing prehistoric socioeconomies from paleoethnobotanical and zooarchaeological data: an example from the British Columbia Plateau. Journal of Ethnobiology 16, 31–62.

Lyman, R.L., 1991. Prehistory of the Oregon Coast. Academic Press, San Diego.

Lyman, R.L., Ames, K.M., 2004. Sampling to redundancy in zooarchaeology: lessons from the Portland Basin, northwestern Oregon and southwestern Washington. Journal of Ethnobiology 24, 329–346.

Lyman, R.L., Wolverton, S., O'Brien, M.J., 1998. Seriation, superposition, and interdigitation: a history of Americanist graphic depictions of culture change. American Antiquity 63, 239–261.

McCartney, P.H., Glass, M.F., 1990. Simulation models and the interpretation of archaeological diversity. American Antiquity 55, 521–536.

Meltzer, D.J., Leonard, R.D., Stratton, S.K., 1992. The relationship between sample size and diversity in archaeological assemblages. Journal of Archaeological Science 19, 375–387.

Monks, G.G., 2000. How much is enough? An approach to sampling ichthyofaunas. Ontario Archaeology 69, 65–75.

Mueller, J.W. (Ed.), 1975. Sampling in Archaeology. University of Arizona Press, Tucson.

Plog, S., Hegmon, M., 1993. The sample size-richness relation: the relevance of research questions, sampling strategies, and behavioral variation. American Antiquity 58, 489–496.

Plog, S., Hegmon, M., 1997. An anthropological perspective on the sample size-richness relation: a response to Leonard. American Antiquity 62, 717–718.

Raup, D.M., 1975. Taxonomic diversity estimation using rarefaction. Paleobiology 1, 333–342.

Rhode, D., 1988. Measurement of archaeological diversity and the sample-size effect. American Antiquity 53, 708–716.

Ringrose, T.J., 1993. Diversity indices and archaeology. In: Andresen, J., Madsen, T., Scollar, I. (Eds.), Computing the Past: CAA92. Aarhus University Press, Aarhus, pp. 279–285.

Sanders, H.L., 1968. Marine benthic diversity: a comparative study. American Naturalist 102, 243–282.

Scheiner, S.M., 2003. Six types of species-area curve. Global Ecology and Biogeography 12, 441–447.

Schoener, T.W., 1986. Patterns in terrestrial vertebrate vs. arthropod communities: do systematic differences in regulation exist? In: Diamond, J., Case, T.J. (Eds.), Community Ecology. Harper and Row, New York, pp. 556–586.

Sharp, N.D., 1990. Fremont and Anasazi resource selection: an examination of faunal assemblage variation in the northern Southwest. The Kiva 56, 45–65.

Shaw, I., 1995. The simulation of artifact diversity at el-Amarna, Egypt. Journal of Field Archaeology 22, 223–238.

Shott, M.J., 1989. Diversity, organization, and behavior in the material record: ethnograhpic and archaeological examples. Current Anthropology 30, 283–315.

Simek, J.F., Snyder, L.M., 1988. Changing assemblage diversity in Periogord archaeofaunas. In: Dibble, H.L., Montet-White, A. (Eds.), Upper Pleistocene Prehistory of Western Eurasia. University of Pennsylvania Press, Pittsburgh, pp. 321–332.

Styles, B.W., 1978. Faunal Exploitation and Energy Expenditure: Early Late Woodland Subsistence in the Lower Illinois Valley. Doctoral dissertation, Northwestern University, Evanston, Illinois.

Styles, B.W., 1981. Faunal Exploitation and Resource Selection: Early Late Woodland Subsistence in the Lower Illinois Valley. Northwestern University Archeological Program, Evanston, Illinois.

Sullivan III, A.P., Tolonen, A.S., 1998. Evaluating assemblage diversity measures with surface archaeological data. In: Sullivan III, A.P. (Ed.), Surface Archaeology. University of New Mexico Press, Albuquerque, pp. 143–155.

Thomas, D.H., 1983. Mid-range theory: sample size effect. In: Thomas, D.H. (Ed.), The Archaeology of Monitor Valley 2: Gatecliff Shelter. Anthropological Papers 59. American Museum of Natural History, New York, pp. 425–433.

Tipper, J.C., 1979. Rarefaction and rarefiction—the use and abuse of a method in paleoecology. Paleobiology 5, 423–434.

Wolff, R.G., 1975. Sampling and sample size in ecological analyses of fossil mammals. Paleobiology 1, 195–204.

Zohar, I., Belmaker, M., 2005. Size does matter: methodological comments on sieve size and species richness in fishbone assemblages. Journal of Archaeological Science 32, 635–641.