

NOTE

QUANTITATIVE METHOD FOR DETERMINING A REPRESENTATIVE ALGAL SAMPLE COUNT¹

Janice L. Pappas² and Eugene F. Stoermer

Center for Great Lakes and Aquatic Sciences, University of Michigan, Ann Arbor, Michigan 48109-2099

ABSTRACT

A method for determining a representative count of a sample dependent on number of species is presented for application to various algal communities. Constant species curves are calculated as $\text{efficiency} = (\text{number of individuals} - \text{number of species}) / \text{number of individuals}$ and diagrammed on a plot of efficiency versus number of individuals counted. Efficiency is defined as the probability that a new species encountered is minimal. That is, as the ratio of number of species to number of individuals approaches 1, more individuals will need to be counted in order to achieve a representative count. Data and calculations of efficiency from two algal communities are presented for illustration.

Key index words: constant species curves; efficiency; probability; specimen counting

Methods of obtaining a representative sample of algal assemblages from various habitats have been an ongoing matter of discussion. Furthermore, preparation of slide mounts and subsequent counting of specimens are also a matter of concern. How many individuals should be counted from a slide mount to represent a sample of an algal assemblage in a given habitat? What is the difference in number of individuals that should be counted to assure adequate representation of different habitats or communities? These broader questions are of interest to algal ecologists.

Representative counts of samples involve a significant investment of time and effort. Generally, maximum number of species in a community is given by the mode of a log normal distribution, when number of species is plotted versus number of individuals in intervals on a logarithmic scale (Patrick et al. 1954, Patrick 1963, Patrick and Strawbridge 1963). However, what is the relation between maximum number of species in numbers of communities and the number of individuals counted in numbers of communities?

Consideration of sample size may be important for comparison of different algal collections from different kinds of communities; that is, collections may be compared if collection area size is the same or each collection consists of a random sample of

individuals (Hurlbert 1971). However, number of species for a single sample size may also indicate an erroneous proportion of number of species to number of individuals (Hurlbert 1971); for example, counting 500 specimens for a benthic community would not necessarily be indicative of proportion of number of species to number of individuals. According to Hurlbert (1971), the most useful approach would be to calculate, plot, and compare species richness curves or plots of a random selection of a fixed number of individuals in a collection versus expected number of species in a sample.

In palynology, pollen grains are classified as categories or types, and counts are viewed as multiple Bernoulli trials. If a specific type(s) of grain(s) is (are) chosen to be counted to a fixed total, various distributions of the counts emerge. Counting to a fixed number for each of three types produces a joint multinomial distribution with each type following a binomial distribution (Mosimann 1965). If two types are counted while only the third type is counted to a fixed number, the two types jointly follow a negative multinomial distribution with each of the two types following a negative binomial distribution. Once the sampling plan is chosen, either the multinomial or negative multinomial distribution applies. Multinomial distribution applies to within the pollen sum, whereas negative multinomial distribution applies to outside the pollen sum (Mosimann 1965).

The assumption of multiple Bernoulli trials means that specimens are independent and that each specimen counted has an equal probability of occurrence. For algal counts, is the distribution of species on a slide truly random, and therefore does each species have an equal probability of occurrence? There are many colonial forms so that the next specimen within a colony is the only real possibility for that trial. In addition, if algal slide preparations reveal clumped rather than random distributions, then the proportion of number of species to number of individuals counted will not be accurate with the assumption of multiple Bernoulli trials.

How does one choose a valid number for a fixed count prior to commencing the count? What are reasonable criteria for number of specimens to be counted? Slide preparations from same-size samples of different algal communities have a wide range in number of species and specimens per slide. Counting a marker specimen to a fixed number may not en-

¹ Received 23 October 1995. Accepted 25 April 1996.

² Author for reprint requests.

sure counting enough of the other specimens to get a representative count of an algal community.

What constitutes a representative count? The problem may be viewed as a matter of efficacy and efficiency. That is, to be most effective, the number of individuals counted representing species composition must be enough such that the probability of encountering a new species is minimal. The result would approximate maximum efficiency in counting.

To study the problem of what constitutes a representative count, Patrick's (1963) notion that number of species versus number of individuals is logarithmic will be the starting point. Rather than using predetermined fixed count totals as Mosimann (1965) suggested, number of individuals may be determined early during the course of counting. Following Hurlbert's (1971) suggestion, we present graphical diagrams useful in determining counts of individuals for number of species found. For illustration, planktonic and benthic communities will be examined. Elucidating a method for counting a representative number of individuals of species composition across many kinds of communities will provide a formal procedure for algal specimen counting and provide confidence in results obtained.

Theory. Except in limiting cases, as counts are extended, number of species discovered is inversely proportional to number of individuals already enumerated. That is, as more individuals are counted, fewer new species will be found. Generally, this relation is represented mathematically as

$$f(x) = a \ln(x) - b. \quad (1)$$

For n species for n communities, equation (1) is modified as

$$f(x) = a \ln(x - n) - b, \quad (2)$$

where a and b are constants, n is the number of species, and x is the number of individuals to be counted. As n is changed in value, separate equations for each of n species are determined, and n constant species curves result. When $f(x) = 0$, $n = x$.

The first derivative, and therefore the slope, is

$$f'(x) = a/x - n, \quad x > n, \quad (3)$$

or

$$f'(x) = a/x_{0+k} - x_0, \quad (4)$$

where x_{0+k} is the number of individuals to be counted, x_0 is the number of species, and k is an integer greater than 0. The slope may also be represented as

$$m = \Delta y / \Delta x \quad (5)$$

or

$$m = (y_{0+k} - y_0) / (x_{0+k} - x_0) \quad (6)$$

and

$$f''(x) = -a/(x - n)^2. \quad (7)$$

From the Mean Value Theorem,

$$(y_{0+k} - y_0) = f'(x)(x_{0+k} - x_0). \quad (8)$$

Therefore,

$$a/(x_{0+k} - x_0) = (y_{0+k} - y_0)/(x_{0+k} - x_0), \quad (9)$$

where x_{0+k} is the number of individuals to be counted.

The ratio of number of species to number of individuals counted results in a numerical value between 0 and 1. Number of species may be defined as number of occurrences, whereas number of individuals counted may be defined as number of trials. Therefore, the ratio of species to number of individuals is a probability value. Because number of species and number of individuals are complementary events, probability may be written as

$$\text{efficiency} = \frac{\text{number of individuals} - \text{number of species}}{\text{number of individuals}} \quad (10)$$

or

$$\text{efficiency} = 1 - \frac{\text{number of species}}{\text{number of individuals}}, \quad (11)$$

where efficiency is the probability that the necessary number of individuals have been counted. Maximum efficiency in counting means that the probability that a new species will be encountered is minimal; that is, the ratio of number of species to number of individuals is nearer to 0 than to 1.

Let $g(x) = \text{efficiency}$. For $a = 1$.

$$1 - x_0/x_{0+k} = a \ln(x_{0+k} - x_0) - b \quad (\text{see below}). \quad (12)$$

Proof of equation (12). We prove that

$$1 - x_0/x_{0+k} = a \ln(x_{0+k} - x_0) - b \quad \text{for } a = 1.$$

The general mathematical form of constant species curves is

$$f(x) = a \ln(x_{0+k} - x_0) - b \quad \text{with } f'(x) = a/x_{0+k} - x_0.$$

From the Mean Value Theorem,

$$x_{0+k} = x_0/1 - y_{0+k} + y_0.$$

Let $g(x) = \text{efficiency}$, which we denote as

$$g(x) = 1 - x_0/x_{0+k} \quad \text{with } g'(x) = (x_0/x_{0+k})^2.$$

From the Mean Value Theorem,

$$x_{0+k} = -x_0/y_{0+k} - y_0 - 1.$$

Therefore, $f(x) = g(x)$.

For illustration, number of species and counts of individuals were available for two different communities. One includes slide data from plankton samples from Lake Huron near shore surface waters (Pappas and Stoermer 1995). The other community is slide data from benthic diatom samples from Quebec Harbor, Michipicoten Island, Lake Superior (Stoermer 1975).

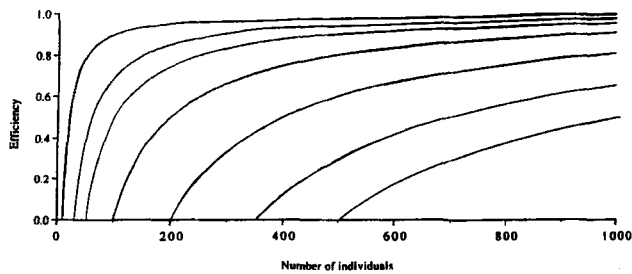


FIG. 1. From left to right, constant species response curves for 10, 30, 50, 100, 200, 350, and 500 species diagrammed on a plot of efficiency versus number of individuals.

Constant species curves were calculated and diagrammed on a plot of efficiency versus number of individuals. Each constant species curve may be used to determine number of individuals to count for number of species on a slide in order to achieve a particular efficiency. As examples, curves for 10, 30, 50, 100, 200, 350, and 500 species are presented (Fig. 1). Extending the ends of the curves to asymptotic efficiency indicates the maximum number to be counted. For the specified number of species, efficiency of number of individuals counted may be determined from Figure 1 or calculated using equation (10) or (11).

When commencing a count, number of species on a slide is usually unknown. However, in aquatic samples, information such as location of collection, depth of collection, offshore or nearshore collection site, habitat of collection site, the time of year of collection, and other factors are useful in determining a first guess as to how many species may be present. By counting a few individuals, say, 10–30, a preliminary check of efficiency may be made. As number of species found approaches number of individuals counted, an increasing number of individuals would need to be counted to achieve high efficiency (Fig. 1). That is, the probability that a new species will be encountered is high; the ratio of number of species to number of individuals is nearer to 1 than to 0. This may be easily checked by using equation (10) or (11) or constructing a constant species curve such as those depicted in Figure 1. For each $n = 1, 2, \dots$, to maximum number of species likely, constant species curves may be calculated and diagrammed. Alternatively, efficiency may be approximated or interpolated from a diagram with only a few species curves (Fig. 1).

Between communities, level of efficiency should be (approximately) the same for ease of comparison. However, if efficiencies are very different, especially if they are lower for a given community, construction of constant species curves may be used to determine the number of individuals that should have been counted to achieve the efficiency being used for comparison. In this way, representative counts of species composition may be ensured at a specified efficiency.

TABLE 1. Number of species, number of individuals, and efficiencies for algal samples from Lake Huron (Pappas and Stoermer 1995) and Lake Superior (Stoermer 1975).

Location	Number of species	Number of individuals	Efficiency	
Lake Huron	39	505	0.923	
	47	842	0.944	
	52	730	0.929	
	54	643	0.916	
	60	856	0.930	
	67	644	0.896	
	67	952	0.930	
	70	826	0.915	
	87	1203	0.928	
	Lake Superior	73	1827	0.960
		84	7961	0.989
		142	1734	0.918
		176	12,774	0.986
		179	19,803	0.991
202		10,558	0.981	
202		18,896	0.989	
204		10,450	0.980	
227		91,696	0.998	
231		42,812	0.995	
232		19,490	0.988	
295	59,151	0.995		
321	63,973	0.995		
339	143,820	0.998		

Illustrations. For each set of data, number of species, number of individuals counted, and the efficiency (calculated from equation (10) or (11)) are given in Table 1. Each row in the table represents individual slide mounts. Lake Huron slides have small numbers of species with number of individuals counted, resulting in 90% or greater efficiency. For the two slides with 67 species each, an increase in count of 308 individuals produced an increase in efficiency of 3.4% (Table 1). From equation (10) or (11) for 67 species, if only 75% efficiency is necessary, only approximately 268 individuals would need to be counted. However, to achieve asymptotic efficiency approximately, 6700 individuals would need to be counted.

Extended counts were performed with Lake Superior slides because number of species ranged from 73 to 339. Therefore, high efficiencies resulted. Except for the slide with the fewest number counted, efficiencies were at least 96% (Table 1). Slides with 202 species each had a difference in count of 8338 with a 0.8% increase efficiency resulting from the larger count. It may be judged that the additional efficiency was not worth the time and effort. When the objective is to obtain a representative count for the number of species present, a particular level of efficiency is the deciding factor, weighing time for additional counting versus a lesser efficiency. For Lake Superior slide counts, approaching asymptotic efficiency (99.8%) for number of species at 227 and 339 meant number of individuals counted was 91,696 and 143,820, respectively. Again, the time necessary to produce such results in contrast to less time count-

ing with a lower efficiency would need to be weighed by each investigator. Depending on how the results of counts are to be used, or if the expectation is that the algal assemblage has a great number of species, approaching asymptotic efficiency in counting may be warranted.

For between community comparison at approximately 92% efficiency, 643 individuals were counted for 54 species, and 826 were counted for 70 species for two Lake Huron slides: 1734 were counted for 142 species for a Lake Superior slide. This shows the progression of number of individuals needed to be counted for the same efficiency. The proportion of number of species to number of individuals remains constant across both communities.

Determining representative counts of species from slides of algal samples has been formalized by the method outlined in this study. Application of this method with efficiency in counting will be useful in

aiding comparisons of species composition and abundances in algal assemblages.

- Hurlbert, S. H. 1971. The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52:577-86.
- Mosimann, J. E. 1965. Statistical methods for the pollen analyst: multinomial and negative multinomial techniques. In Kummel, B. & Raup, D. [Eds.] *Handbook of Paleontological Techniques*. W. H. Freeman, San Francisco, pp. 633-73.
- Pappas, J. L. & Stoermer, E. F. 1995. Effects of inorganic nitrogen on Lake Huron phytoplankton: an experimental study. *J. Great Lakes Res.* 21:178-91.
- Patrick, R. 1963. The structure of diatom communities under varying ecological conditions. *Ann. N.Y. Acad. Sci.* 108:353-8.
- Patrick, R., Hohn, M. H. & Wallace, J. H. 1954. A new method for determining the pattern of the diatom flora. *Not. Nat.* 259:1-12.
- Patrick, R. & Strawbridge, D. 1963. Variation in the structure of natural diatom communities. *Am. Nat.* 97:51-7.
- Stoermer, E. F. 1975. Comparison of benthic diatom communities in Lake Michigan and Lake Superior. *Verh. Int. Verein. Limnol.* 19:932-8.