Use of a Statistical Technique for Recommending Sample Size in Collecting Benthic Macroinvertebrate Populations[1]

JERRY L. WILHM, Oklahoma State University, Stillwater

Determination of the number of replicate samples necessary to obtain an adequate representation of species present in a particular habitat is often a major problem in studies of benthic macroinvertebrate populations. Gaufin, Harris and Walter (1956) described a statistical evaluation of stream-bottom sampling data obtained from three standard samplers, a long handled dip strainer, an Ekman dredge, and a Surber square-foot sampler. In the present study, their technique was used in pilot studies of a polluted and a nonpolluted habitat. Large numbers of selected samples of benthic macroinvertebrates were taken and recommendations were suggested for some smaller number of samples.

This technique provides a statistical criterion for determining the proportion of species present in a particular habitat that would be detected on the average after collecting a certain number of samples. Initially, a predetermined number of replicate samples assumed to contain 100% of the species must be collected. While the number of samples required to collect all species in a nonpolluted environment may be extremely large, this technique can be used adequately with a smaller number of samples. For the present study, $n = 10$ was selected.

A coefficients table (Table I) was constructed by the formula,

$$a_{ik} = ({}_{n-k+1}C_i)\ (i)\ /\ ({}_nC_i)\ (n-k+1)$$

where a_{ik} is the coefficient in the ith row and kth column, n is the sample size, and ${}_{n-k+1}C_i$ and ${}_nC_i$ are combinations,

$${}_nC_i = (n!)\ /\ (i!)\ (n-i)!.$$

Row i provides a distribution of probabilities that a species having appeared i times will have made its initial appearance in the kth sample. For example, if a species appears in all ten samples ($i = 10$) the probability that it will appear initially in the first sample ($k = 1$) is 1 and a_{10} is the only coefficient required in row 10. A species appearing in five samples ($i = 5$) has a probability of 0.500 of appearing initially in the first sample ($k = 1$) and no probability of appearing initially in the seventh

TABLE I. COEFFICIENTS TABLE

$$a_{ik} = ({}_{n-k+1}C_i)\ (i)\ /\ ({}_nC_i)\ (n-k+1)$$

	k									
i	1	2	3	4	5	6	7	8	9	10
1	.100	.100	.100	.100	.100	.100	.100	.100	.100	.100
2	.200	.178	.156	.133	.111	.089	.067	.044	.022	
3	.300	.233	.175	.125	.083	.050	.025	.008		
4	.400	.267	.167	.095	.048	.019	.005			
5	.500	.278	.139	.060	.020	.004				
6	.600	.227	.100	.028	.005					
7	.700	.233	.058	.008						
8	.800	.178	.022							
9	.900	.100								
10	1.000									

Vacant half of table is an area of zero probability.

[1]Contribution No. 434 from Zoology Department, Oklahoma State University.

sample. Finally, if a species appears in only one sample ($i = 1$), a probability of 0.100 exists that it will appear initially in any of the ten samples.

Field data can be applied to this table by the formula,

$$P_k = \sum_{i=1}^{n-k+1} a_{ik} \frac{m_i}{m}$$

where P_k is the average probability of species appearing initially in the kth sample, m is the total number of species collected in n samples, and m_i is the number of species appearing in i samples.

This technique was applied to samples of benthic macroinvertebrates from polluted and nonpolluted areas of Skeleton Creek, Oklahoma, and from a spring in the headwaters of White Oak Creek, Tennessee. Skeleton Creek is a permanent stream which originates near Enid, Oklahoma, flows southeasterly for 70 miles, and empties into the Cimarron River 5 miles north of Guthrie, Oklahoma. Both municipal and industrial wastes enter the headwaters of Skeleton Creek. Stations in Skeleton Creek were 11, 25, 43 and 60 miles below the source of the effluent. Stream characteristics and benthic macroinvertebrate populations were described by Wilhm and Dorris (1966). The spring area is located in Roane County, Tennessee. The area of the basin is 93 m^2 and depth increases from several centimeters near the margin to 0.6 m in the middle. Most of the bottom is covered with mud. Discharge averages 0.0028 m^3/sec.

Ten replicate Ekman dredge samples were taken from marginal areas at four different stations in Skeleton Creek on 13 February 1965 and from the spring on 7 April 1965. Samples were washed in a sieve (#40 U.S. standard soil series with openings of 0.420 mm) and preserved in alcohol.

Distribution of species according to the number of samples in which each species appears is given in Table II and cumulative P_k by stations in Table III. In Skeleton Creek, numbers of species varied from five at the uppermost stations to 19 at the lowermost station (Table II). Two species were found in all ten samples at Station 11, one at Station 25, and none at Stations 43 or 60. One sample yielded on the average half of the species observed after ten samples at Station 11 and less than 30% at Stations 43 and 60 (Table III). Three samples at the lowermost station were required to produce 50% of the species observed after ten samples.

Twenty-one species were collected from the spring located in the headwaters of White Oak Creek (Table II). Seven species were present in only one sample and no species were represented in all ten samples. Two samples yielded almost half of the species observed aften ten samples and 75% of the species were not observed until the sixth sample.

The large numbers of selected samples obtained from Skeleton Creek were used for recommendation of sample size in the investigation of community structure of benthic macroinvertebrates. Data obtained from the spring were used in a pilot study preceding the investigation of community organization and energy flow of benthic macroinvertebrates. The small size of the spring and the labor and time involved in collecting and identifying organisms in both the stream and the spring prevented a large number of replications. Thus, sampling had to be adequate but not redundant. Collection of data and evaluation of cumulative P_k provided useful information in determining the number of replicate samples to collect.

TABLE II. DISTRIBUTION OF SPECIES (m_i) ACCORDING TO THE NUMBER OF SAMPLES (i) IN WHICH EACH SPECIES APPEARS.

	Skeleton Creek				White Oak
i	11	25	43	60	Spring
1	2	2	7	7	7
2	0	3	4	4	4
3	1	0	0	3	4
4	0	2	0	2	1
5	0	0	1	1	2
6	0	0	0	0	2
7	0	1	2	1	1
8	0	0	0	1	0
9	0	0	1	0	0
10	2	1	0	0	0
Total (m)	5	9	15	19	21

TABLE III. CUMULATIVE P_k BY STATIONS

$$P_k = \sum_{i=1}^{n-k+1} a_{ik} \frac{m_i}{m}$$

	Skeleton Creek				White Oak
k	11	25	43	60	Spring
1	.500	.367	.287	.274	.319
2	.587	.534	.437	.450	.498
3	.662	.652	.542	.576	.620
4	.727	.740	.629	.674	.718
5	.784	.810	.707	.753	.787
6	.734	.866	.778	.812	.843
7	.879	.912	.842	.874	.890
8	.920	.949	.900	.921	.931
9	.960	.974	.953	.963	.967
10	1.000	1.000	1.000	1.000	1.000

LITERATURE CITED

Gaufin, A. R., E. K. Harris and H. J. Walter. 1956. A statistical evaluation of stream bottom sampling data obtained from three standard samplers. Ecology 37: 643-48.

Wilhm, J. L. and T. C. Dorris. 1966. Species diversity of benthic macroinvertebrates in a stream receiving domestic and oil refinery effluents. Amer. Midl. Nat. 76:427-49.