Sampling Design and Enumeration Statistics for Bacteria Extracted from Marine Sediments†

PAUL A. MONTAGNA

Belle W. Baruch Institute for Marine Biology and Coastal Research and Department of Biology, University of South Carolina, Columbia, South Carolina 29208

The spatial and temporal distributions of marine bacteria were studied at both a muddy and a sandy subtidal site in North Inlet, S.C. The sampling design was hierarchical, since subsampling (by a dilution series) of the sediments was necessary to count bacterial cells using acridine orange epifluoresence microscopy. The cell count data fit a log-normal distribution. The abundance of bacteria was 10^{11} g^{-1} (dry weight) of mud and 10^9 g^{-1} (dry weight) of sand. Variance component analyses demonstrated that variation due to the subsampling procedures was always statistically significant. Thus the common practice of counting 20 fields from one filter preparation is inadequate for estimating the true bacterial population variance in marine sediments. It is recommended that replication of the subsampling level be performed. Standardization of data (by dry weight of sediment) decreased sampling variance at the mud site but not at the sand site, implying that bacteria are more homogeneously distributed in sand than in mud.

There is a long history of bacterial counts in the water column dating back to the period between 1900 and 1930 (9, 11). However, abundance estimates of bacteria in marine sediments have also become a subject of much interest (22). Direct count data of cells from marine sediments are becoming more and more common with the recent advent and widespread use of epifluorescence microscopy (6, 16).

Common statistical techniques used with bacterial count data assume either a Poisson or a negative binomial distribution (8, 17). These statistical distributions have been useful in quantifying the number of coliforms in the water, thereby measuring water quality. In Poisson distributions coliform bacteria are assumed to be randomly distributed in small-volume samples (11). Since the variance equals the mean, Poisson distributions are easy to work with because only one parameter of the population must be measured.

Fisher et al. (9) found that Poisson distributions fit bacteria populations under carefully controlled experimental conditions. However, natural populations in lake water samples fit negative binomial distributions but do not fit Poisson distributions (8, 17). In negative binomial distributions the variance is greater than the mean because of additional variation due to departures from randomness. Both distributions are in the binomial family and are useful in

describing count data generated in sampling bacteria such as coliforms. However, it is doubtful that naturally occurring bacteria in marine sediments are distributed in the same fashion as coliforms in the water column. Bacteria in sediments are numerous, require dilution for counting, and adhere to surfaces of particles; thus, their distribution may not be random. In this paper I examine the statistical aspects of counting bacterial cells in two common types of marine sediments.

MATERIALS AND METHODS

The study area was located in the North Inlet estuary, Georgetown, S.C. (33° 20.0′N, 79° 10.0′W). The two study sites, one sand and one mud, have been described by Coull and Fleeger (5). At both sites, sediment cores were taken at a water depth of 40 cm during low tide with a 3-cm^3 syringe barrel with the bottom removed (inner diameter = 8.5 mm) to a sediment depth of 8.7 mm; thus the total area sampled was 56.7 mm^2, and the total volume was 0.5 cm^3. Sediments were fixed in 4% buffered Formalin and analyzed within 2 weeks. Mud sediments were diluted 10,000 times, and sand was diluted 400 times to obtain approximately 50 cells per microscope field. In addition, it was necessary for sand samples to settle for 30 s to remove large particles; this did not affect abundance estimates. Sterile distilled water was used for dilutions because homogenized seawater foamed, making pipetting difficult.

Sediment presents three problems to direct count techniques: (i) bacteria may be underneath the particles or otherwise out of view, (ii) associated detritus produces a background fluorescence making identification of bacteria difficult, (iii) particles (particularly
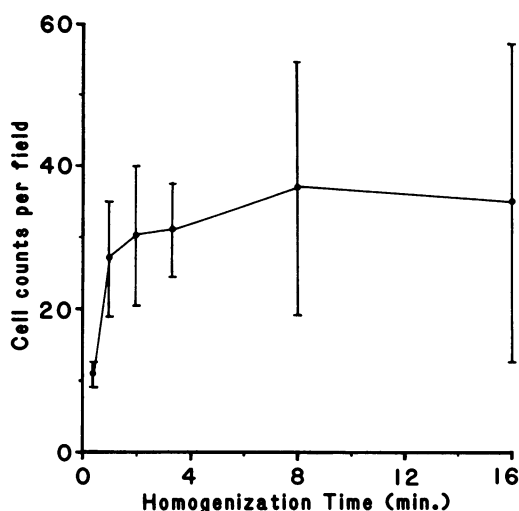
FIG. 1. Number of cell counts per field versus the homogenization time of a mud sample. Mean of five microscope fields and the 95% confidence intervals are given.

sand) increase the working distance of the microscope objective, distorting optical quality. Ideally, the bacteria are completely separated from the sediment. Homogenization is the best dispersion technique (6, 16); however, underestimates could be expected due to cell lysis and the inability to remove all bacteria from the sediments. Homogenization for more than 30 s yielded consistent results (Fig. 1), contrast method, $P = 0.0004$ (12); the means are equal, Duncan's multiple range test, $P > 0.05$ (12); and the variances are homogeneous, Bartlett's test, $P > 0.05$ (19). I chose to homogenize sediments for 10 min (although 4 would have been adequate; Fig. 1) with a Waring commercial blender (23,000 rpm). The major advantage for blending at least 4 min is that the detritus particles are reduced sufficiently so that background fluorescence and particle interference are eliminated.

Acridine orange epifluorescence microscopy were used to obtain direct cell counts from the diluted samples (7, 13). Nuclepore filters (0.2-μm pore diameter, 25-mm diameter) prestained with Irgalin black were used (3): 1 ml of diluted sample was incubated for 4 min with 1 ml of 0.05% acridine orange, and the filter was washed with 2 ml of sterile distilled water. Slides of the filters were examined at ×1,250, using a Leitz Dialμx 20 fitted with a Ploempak 2.4 system, HBO 50W burner, and a 475 edge filter.

Standardization of the data was achieved by filtering 20 ml of the homogenized sample and measuring the dry weight of the sediment contained in each core sample. The samples were dried at 55°C for 24 h.

Statistical considerations and sampling design. It is common practice when measuring bacterial abundances in marine sediments to count between 20 and 25 microscope fields for each filter preparation, a practice borrowed from water column studies (6, 10, 18). Assuming a Poisson (4) or a normal distribution (14), there is a linear relationship between confidence limits about the mean and number of fields counted.

From this relationship 20 fields would yield deviations of only ±20% when the mean is greater than 30 cell counts per field (14). There is a negative hyperbolic relationship between the mean and the number of fields required; thus many more fields are required to achieve the same precision when the mean is small, i.e., <20 (14). In sediments the mean is typically large.

Water column workers generally filter a fixed volume of water, but analyzing sediments requires dilution and subsampling. The variance of the population is no longer a result of the samples alone, since subsampling introduces variability (15, 21). Therefore it is necessary also to measure subsampling variance to distinguish between samples.

The first experiment was designed to measure population variance and the most efficient ways to estimate them. A completely balanced and randomized design was employed, using 4 sediment cores with 4 filters per core and 10 microscope fields per filter, making 160 cell counts per core (or sample). Subsampling schemes such as this are analogous to nested or hierarchical sampling plans (21). In this design the independent variable was counts of bacterial cells per field. The dependent variables were the filter that was counted and the core from which the filter came (a nested factor). The statistical model for the counts of bacterial cells per field was $y_{ijk} = \mu + C_i + F_{ij} + e_{ijk}$, where μ is the overall sample mean, C_i is the counts per core, F_{ij} is the counts per filter per core, and e_{ijk} is the random error due to counts from each microscope field. The model for the variance (σ^2) was $\sigma_y^2 = \sigma_C^2 + \sigma_F^2 + \sigma^2$. The optimal sampling design suggested from the first experiment was used in the second experiment to compare the mud and sand stations over three different sampling periods. All data analyses were performed with Statistical Analysis System software (12).

RESULTS

Sampling design experiment. The cell count data did not fit a normal distribution (Kolmogorov-Smirnov test, $n = 160$, $P < 0.01$), but did fit a log-normal distribution because (i) the variance of $\log_{10}$-transformed counts are independent of the mean ($r^2 = 0.056$, $P = 0.377$ for the null hypothesis that the slope equals zero); (ii) the data fit a log-normal distribution in the Kolmogorov-Smirnov test ($P = 0.139$). There was no evidence that the data fit a Poisson distribution (chi-square goodness-of-fit test, $P < 0.0001$) or a negative binomial distribution (chi-square goodness-of-fit test, $P < 0.0001$).

By knowing the underlying distribution of the population and using the appropriate transformation to normalize the data, one can confidently use analysis of variance procedures to analyze the data. The only further, but necessary, assumption is additivity of the variances (see above).

Analysis of the sampling design experiment indicated significant differences between cores and between filters within cores (Table 1). Subsampling (i.e., the dilution process) added vari-

TABLE 1. Hierarchical analysis of variance for the sampling design experiment at the mud site[a]

Source	df	Mean square	Expected MS	F	P	% Variance
Core	3	0.8176781	$\sigma^2 + \sigma_F^2 + \sigma_C^2$	7.12	0.0053	28.7
Filter	12	0.1148754	$\sigma^2 + \sigma_F^2$	3.21	0.0004	12.9
Field	144	0.0357495	σ^2			58.3

[a] Log-transformed counts (= $\log_{10} x$); coefficient of variation = 10.9%. Abbreviations: MS, mean square; F, field; C, core.

TABLE 2. Hierarchical analysis of variance for the sampling design experiment[a]

Source	df	Mean square	Expected MS	F	P	% Variance
Core	3	0.0561374	$\sigma^2 + \sigma_F^2 + \sigma_C^2$	0.49	0.6966	0.0
Filter	12	0.1148754	$\sigma^2 + \sigma_F^2$	3.21	0.0004	18.1
Field	144	0.0357495	σ^2			81.9

[a] Log-standardized counts, i.e., $\log_{10}$ (x/g [dry weight] of sediment); Coefficient of variation = 7.1%. For abbreviations, see Table 1, footnote a.

ance representing 12.9% of the total. However, since subsampling error was less than the variance among cores (28.7%), significant spatial variation in bacterial counts on the spatial scale of cores (0.57 cm²) could be identified.

Differences among cores can be examined by standardization, i.e., by examining the variance of each core on a sediment-dry-weight basis.
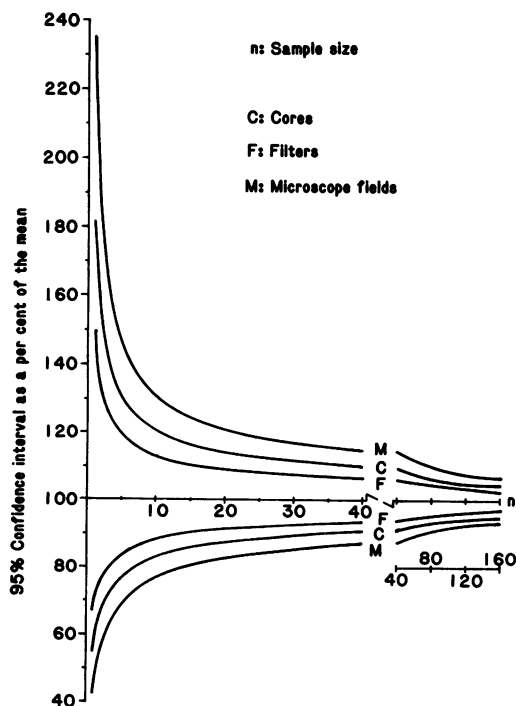

FIG. 2. Diagram for determining the number of samples from each variance component necessary to obtain 95% confidence intervals within a desired percentage of the mean.

When this was done, cores were no longer significantly different and had no variance component (Table 2). All variance was due to the subsampling and counting procedures. Overall variance was also decreased; the coefficient of variation decreased from 10.9 to 7.1%. Apparently equal volumes of mud did not contain equal dry weights of sediments.

The relationship between the number of samples (n) and the percent change in the 95% confidence interval as a percentage of the mean is given by

$$\overline{X} \times/\div \frac{\text{antilog} [t \times (\text{MS}/n)^{1/2}]}{\overline{X}} \times 100\%$$

where $\times/\div$ indicates multiplication or division, t is 1.96 (the two-tailed 95% confidence limit with 159 df) and MS is the mean square of the variance component. This relationship is illustrated in Fig. 2 and is asymetric because it was derived from detransformed log data. It is most appropriate at the mud site and is dependent on having approximately 50 cells per field. However, the basic relationship is consistent with results from the sand site. It indicates that accuracy can be achieved by counting 7 to 11 cores, 3 to 5 filters, and 15 to 21 microscope fields. When the data were standardized, the variance component due to cores became essentially zero (Table 2), and the line C in Fig. 2 disappeared.

From the results of the above study, an optimal sampling design can be found which is less time consuming and more efficient for estimating the mean and variance of the population. The expected value of the variance is given by Sokal and Rolf (19) as

$$s^2 = \frac{s^2_{(core)}}{i} + \frac{s^2_{(filter)}}{ij} + \frac{s^2_{(field)}}{ijk}$$

TABLE 3. Comparison of possible experimental designs with the sampling design experiment

Source	Log-transformed data from design experiment		No. of replicates from possible design:			
	Variance estimate	No. of replicates	I	II	III	IV
Cores	0.01757007	4	1	2	2	2
Filters	0.00791259	4	1	2	2	2
Fields	0.03574953	10	20	5	10	15
Total fields counted	160		20	20	40	60
Total variance estimate	0.0051105		0.0272702	0.0125507	0.0116569	0.0113590
% Increase in variance estimate from experiment			434	146	128	122

where $s^2_{(x)}$ is the variance component of x, $i =$ number of cores, $j =$ number of filters, and $k =$ number of fields. For dry weight standardized data (as in Table 2), the first term (the variance due to the core) would be zero; thus no replication in that level would be necessary (i would equal unity). However, spatial distribution is a useful characteristic in many studies, so I used the data extracted from the raw counts (i.e., counts per core) in this analysis.

If 20 fields were counted from only one core and one filter, the estimate of the variance would be 217% greater than if five fields were counted from two cores with two filters each (Table 3, examples I and II). By increasing the numbers of fields counted in the former replicate design, variance estimates would be decreased by only 18% if 10 fields per filter were counted (plan III) and by 6% if 15 fields were counted (plan IV). Clearly, replication is most important at the higher levels, such as cores and filters.

Comparisons between stations and sampling periods. From the results of the previous section I decided to use sampling plan IV (15 fields from each of two filters and two cores) from Table 3. Although design II (five fields per filter) would have been only 24% less efficient with much less work, by increasing the total counts from 20 to 60 fields, precision is almost doubled (Fig. 2).

The results of the first experiment were re-peatable during the next two sampling periods (each 2 weeks apart). A high percentage of the variance was found among cores for the count data (Table 4). Although one date was not significant, variance among cores could be decreased by standardization (Table 5). In either treatment of the data, variance due to subsampling (filters) was always significant.

The trends at the sand site were not as clear. Generally a high (but not significantly different) proportion of variance was found among cores in both count (Table 4) and standardized data (Table 5). Although subsampling (i.e., filters) was significant on the first sampling date, it was not on the next two dates, and standardization had no effect (Tables 4 and 5).

Although the above treatment of the data (by site and date individually) is useful in comparing the repeatability of sampling experiments, it is more appropriate to analyze the entire experiment with one analysis of variance. Since the abundance of bacteria at the mud station was always two orders of magnitude greater than that of the sand station (Table 6), I treated the two sites independently. Thus, the new model, including sampling periods, is a completely nested design, filters within cores and cores within dates. At the mud station there was no significant difference among dates when counts per core were examined, but there was a difference

TABLE 4. Variance components of the temporal and site experiment[a]

Source	% Variance at:					
	Mud station on:			Sand station on:		
	2/5	2/17	3/4	2/5	2/17	3/4
Cores	28.7**	88.0*	53.6ns	60.6ns	40.2ns	55.2*
Filters	12.9***	2.4*	28.5***	18.2***	1.9ns	0.0ns
Fields	58.4	9.6	17.9	21.2	57.9	44.8

[a] Log-transformed counts from a hierarchical design. Table gives percent variance and significance level (ns = not significant, * = significant at 0.05 level, ** = significant at 0.01 level, and *** = significant at 0.001 level).

TABLE 5. Variance components of the temporal and site experiment[a]

| Source | % Variance at: | | | | | |
| | Mud station on: | | | Sand station on: | | |
	2/5	2/17	3/4	2/5	2/17	3/4
Cores	0.0^{ns}	48.1^{ns}	0.0^{ns}	75.2^{ns}	66.6*	4.0^{ns}
Filters	18.1***	10.5*	61.5***	11.5***	1.0^{ns}	0.0^{ns}
Fields	81.9	41.4	38.5	13.3	32.4	96.0

[a] Log-standardized counts from a hierarchical design. Table gives percent variance and significance level from an F-test (ns = not significant, * = significant at 0.05 level, ** = significant at 0.01 level, and *** = significant at 0.001 level).

when the data were standardized. The variance between cores was too large to distinguish differences among dates. But, as mentioned above, standardization decreased variance between cores, thereby enabling discrimination among dates. In either case, subsampling (within filter variance) was always highly significant (Table 6).

In contrast, standardization of the sand station data did not change the conclusions drawn from the analysis of the data (Table 6). Differences among dates were not significant in either case; differences between cores were significant in both cases. As in the mud station, subsampling was always a significant effect.

DISCUSSION

Sampling aquatic sediments is not like sampling the water column. Bacterial abundances are much greater, and the samples have a large amount of detrital particles which interfere with accurate cell counting. This necessitates the dilution and homogenization procedures which I and other authors have employed (6, 16). From Fig. 1 it appears that subsample variance might be larger at long blending times. This may be due to an increasingly larger number of particles with continued blending that could lead to nonrandom distributions (15). It is evident that these

procedures add to the natural variability found in the field, and this confounds attempts to elucidate the spatial and temporal distributions of natural populations. Variability occurs on at least two levels: variability in the dilution and subsampling process, which may be reasonably assumed to be random, and the variation which exists in the environment. It is the latter natural variability which is usually of interest.

Sediment characteristics play a very important role in structuring natural distributions of the bacteria. There are two orders of magnitude more bacteria in the subtidal muds than in the sandy subtidal sediments of the estuary. Dale (6) also reported a significant negative correlation between sediment grain size and bacterial adherence and suggested the greater surface area of the smaller particles as the controlling factor.

The abundances I found were slightly higher than those reported by Dale (6). In the mud I found 1.3 ($\pm$1.4) $\times$ 10^{11} g^{-1}, whereas Dale, who also used acridine orange techniques, found 1.7 $\times$ 10^9 to 9.9 $\times$ 10^9 g^{-1} in an intertidal basin in Nova Scotia. In the sandy sediments I found 1.3 ($\pm$1.5) $\times$ 10^9 g^{-1}, and Dale (6) found 1.5 $\times$ 10^8 to 3.8 to 10^8 g^{-1}. However, my estimates agree with those of other authors; Meyer-Reil et al. (16) found 0.42 $\times$ 10^9 to 2.72 $\times$ 10^9 g^{-1} in sandy beach sediment from the Baltic Sea, also using

TABLE 6. Comparison of the two sites over the entire experimental period[a]

| Source | % Variance at: | |
	Mud station	Sand station
Log_{10} count core^{-1} ($\bar{X} \pm SD$)	(10.1859 $\pm$ 0.1573)	(8.4069 $\pm$ 0.1796)
Dates	1.9^{ns}	0.0^{ns}
Cores	51.9***	53.9*
Filters	11.4***	10.6***
Fields	34.8	35.5
Log_{10} count g^{-1} ($\bar{X} \pm SD$)	(11.1133 $\pm$ 0.1573)	(9.1144 $\pm$ 0.1796)
Dates	28.8*	0.0^{ns}
Cores	1.6^{ns}	68.4**
Filters	17.2***	7.3***
Fields	52.4	24.3

[a] Table gives percent variance of each hierarchical component and significance of the appropriate F-test (ns = not significant, * = significant at 0.05 level, ** = significant at 0.01 level, and *** = significant at 0.001 level).

acridine orange staining. Anderson and Meadows (1) found 0.14×10^9 to 1.18×10^9 g^{-1} in sandy sediments with phase-contrast microscopy.

The abundance differences within areas related to sediment types are real, but the slight differences due to geographic location may not be. Anderson and Meadows (1) found no differences in abundance associated with tidal levels. But Boeyé et al. (2), using plating techniques, found higher abundances of colony-forming units in nearshore sediments than offshore. Both Anderson and Meadows (1) and Dale (6) report that the surface area of sediment particles have a positive correlation with abundance.

Although I did not find temporal differences over the 6-week study period, Wilson and Stevenson (20) reported that seston bacterial abundance varied tidally and seasonally in the same estuary. Boeyé et al. (2) also found seasonal differences in bacterial abundances of offshore sediments. The lack of temporal differences in this study was probably due to the short time span.

Standardization on a dry weight basis had no effect on abundance estimates at the sand site. In the sandy environment any sample of a given volume, weight, or area can be expected to have the same number of bacterial cells. Replicate volumes or areas in the mud will contain differing amounts of bacteria, but the same dry weights of sediments will contain the same number. This could be due to differences in the variation of compaction, pore water content, sorting, or bioturbation. Thus, any given volumes of muddy sediments could contain different amounts or weights of sediments.

The exact kind of sampling plan an investigator uses to count sediment bacteria would depend on what is being examined (e.g., spatial or temporal distributions). For example if I wished to study temporal variability, one core per visit would be sufficient at the mud site if I standardized the data.

Standardization had no effect on the sand site (Table 6); thus more than one core would be necessary regardless of how I chose to report the data. Spatial differences at either site could be discerned only if cores were replicated. As mentioned above, spatial studies of any scale would also require knowledge of sediment characteristics.

Regardless of what aspect I might choose to study, replication of filters would be very important in assessing the true population variance. For example, I could estimate the variance much more accurately by counting only five fields from each of four filters, rather than 20 fields from one filter preparation, as is common practice.

Presampling studies are necessary to determine the variance of local populations (and the investigator's technique), and where replication should be performed to adequately estimate population variance. This study demonstrates that sampling bacteria in marine sediments is a form of hierarchical sampling and that, in general, replication at the higher rather than lower levels is most appropriate for reducing total variability about an estimate of bacterial abundance.

LITERATURE CITED

1. Anderson, J. G., and P. S. Meadows. 1969. Bacteria on intertidal sand grains. Hydrobiologia 33:33–45.
2. Boeyé, M., M. Wayenbergh, and M. Aerts. 1975. Density and composition of heterotrophic bacterial populations in North Sea sediments. Mar. Biol. 32:263–270.
3. Bowden, W. B. 1977. Comparison of two direct-count techniques for enumerating aquatic bacteria. Appl. Environ. Microbiol. 33:1229–1232.
4. Cassell, E. A. 1965. Rapid graphical method for estimating the precision of direct microscopic counting data. Appl. Environ. Microbiol. 13:293–296.
5. Coull, B. C., and J. W. Fleeger. 1977. Long-term temporal variation and community dynamics of meiobenthic copepods. Ecology 58:1136–1143.
6. Dale, N. G. 1974. Bacteria in intertidal sediments: factors related to their distribution. Limnol. Oceanogr. 19:509–518.
7. Daley, R. J., and J. E. Hobbie. 1975. Direct counts of aquatic bacteria by a modified epifluorescence technique. Limnol. Oceanogr. 20:875–882.
8. El-Shaarawi, A. H., S. R. Esterby, and B. J. Dutka. 1981. Bacterial density in water determined by Poisson or negative binomial distributions. Appl. Environ. Microbiol. 41:107–116.
9. Fisher, R. A., H. G. Thornton, and W. A. MacKenzie. 1922. The accuracy of the plating method of estimating the density of bacterial populations. Ann. Appl. Biol. 9:325–359.
10. Fliermans, C. B., and E. L. Schmidt. 1975. Fluorescence microscopy: direct detection, enumeration and spatial distribution of bacteria in aquatic systems. Arch. Hydrobiol. 76:33–42.
11. Greenwood, M., and G. U. Yule. 1917. On the statistical interpretation of some bacteriological methods employed in water analysis. J. Hyg. 16:36–54.
12. Helwig, J. T., and K. A. Council (ed.). 1979. SAS user's guide, 1979 edition. SAS Institute, Inc., Carey, N.C.
13. Hobbie, J. E., R. J. Daley, and S. Jasper. 1977. Use of Nuclepore filters for counting bacteria by fluorescence microscopy. Appl. Environ. Microbiol. 33:1225–1228.
14. Lebedeva, M. N., and G. V. Shumakova. 1969. Reliability of data obtained by the method of a direct count on filters. Mikrobiologiya. 38:293–298. (Translated from Russian).
15. Lund, J. W. G., C. Kipling, and E. D. LeCren. 1958. The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. Hydrobiologia 11:143–170.

16. **Meyer-Reil, L.-A., R. Dawson, G. Liebezeit, and H. Tiedge.** 1978. Fluctuations and interactions of bacterial activity in sandy beach sediments and overlying waters. Mar. Biol. **48:**161–171.

17. **Pipes, W. O., P. Ward, and S. H. Ahn.** 1977. Frequency distributions for coliform bacteria in water. J. Am. Water Works Assoc. **69:**664–668.

18. **Scheraga, M., M. Meskill, and C. D. Litchfield.** 1979. Analysis of methods for the quantitative recovery of bacteria sorbed onto marine sediments, p. 21–39. *In* C. D. Litchfield and P. L. Seyfried (ed.), Methodology of biomass determinations and microbial activities in sedi-

ments. ASTM STP 673. American Society for Testing and Materials, Philadelphia.

19. **Sokal, R. R., and F. J. Rolf.** 1981. Biometry, 2nd ed. W. H. Freeman and Co., San Francisco, Calif.

20. **Wilson, C. A., and L. H. Stevenson.** 1980. The dynamics of the bacterial population associated with a salt marsh. J. Exp. Mar. Biol. Ecol. **48:**123–138.

21. **Venrick, E. L.** 1971. The statistics of subsampling. Limnol. Oceanogr. **16:**811–818.

22. **Zobell, C. E., and C. B. Feltham.** 1942. The bacterial flora of a marine mud flat as an ecological factor. Ecology **23:**69–78.