

Taxonomic level and sample size sufficient for assessing pollution impacts on the Southern California Bight macrobenthos

Steven P. Ferraro, Faith A. Cole

US Environmental Protection Agency, Pacific Ecosystems Branch, Mark O. Hatfield Marine Science Center, Newport, Oregon 97365, USA

ABSTRACT: Macrobenthic data from samples taken in 1980, 1983 and 1985 along a pollution gradient in the Southern California Bight (USA) were analyzed at 5 taxonomic levels (species, genus, family, order, phylum) to determine the taxon and sample size sufficient for assessing pollution impacts on 5 measures of community structure. Two statistical designs were compared: a *t*-test for differences between reference and impacted stations where the error term was (1) among-year variation at the reference station (*impact effects design*), (2) replicate (within-station) sampling error (*location effects design*). The estimated statistical power ($1-\beta$) to detect impacts was a function of type and magnitude of impact, level of taxonomic identification, the statistical design, and the sample size (n_i = number of sampling years at the reference station for the *impact effects design*, and n_l = number of replicate samples per station for the *location effects design*). Four replicate 0.1 m² van Veen grabs per station were needed to ensure community-wide, unbiased estimates of Shannon's, 1–Simpson's and McIntosh's Index. Family-level identification appeared to be a good choice for assessing pollution impacts at the study site as it ensured a high probability ($1-\beta \geq 0.80$) of detecting intermediate or larger impacts on most (*impact effects design*) or all (*location effects design*) of 5 measures of community structure when n_i and $n_l \geq 4$. The level of taxonomic identification and sample size should be considered along with other sampling variables (e.g. sample unit size, sieve mesh size) when seeking a statistically rigorous, cost-effective study design sufficient to meet pollution assessment objectives.

INTRODUCTION

Identification to species, or lowest possible taxon, is common practice in synecology. Species-level identification, however, may not always be necessary in applied studies such as pollution assessment and monitoring (Edwards et al. 1975, Hellawell 1977, Green 1979, McIntyre et al. 1984, Kingston & Riddle 1989). Identifying organisms to the taxonomic level necessary and sufficient to meet a study's objective(s) has been called 'taxonomic sufficiency' (Ellis 1985). As taxonomic sufficiency moves to progressively higher taxa, costs, in terms of the expertise and time needed to identify organisms, decrease. Sampling costs are also, obviously, a function of sample size.

Resh & Unzicker (1975) noted that many macrobenthic invertebrate congeners have wide pollution tolerances and concluded that species-level identification was, therefore, essential for water-quality monitoring. Their conclusion is correct, however, only if the same genera (or higher taxa) with wide tolerances are

dominant in both polluted and unpolluted areas. If this is the usual case, the concept of taxonomic sufficiency will have little practical application in pollution ecology. However, if the abundance and composition of genera (or higher taxa) differ in polluted and unpolluted areas, little or no relevant information may be lost by identifying animals to higher taxa.

The level of taxonomic identification may influence community structure measures often used to assess pollution impacts. Caution is advised when comparing diversity indices based on different taxonomic groupings (Hughes 1978, Wu 1982). The fact that a species diversity index can differ numerically from a supra-specific diversity index, however, does not necessarily negate the usefulness of the latter for discriminating differences between communities. One can estimate the power ($1-\hat{\beta}$, where β equals the probability of Type II error) to detect differences between statistics, including diversity indices, by a *t*-test given the probability of Type I error (α), the sample size (n), and the magnitude of the difference between the statistics in standard

deviation (SD) units (the 'effect size' = $ES = |\bar{X}_1 - \bar{X}_2| / SD$) (Cohen 1977). For constant α and n , any increase in ES , either by an increase in the magnitude of the difference or a decrease in the variance, will increase $1 - \beta$.

Warwick (1988a, b) and Heip et al. (1988) provide the best evidence to date for the taxonomic sufficiency of higher taxa in pollution studies. These studies demonstrated the equivalency of pollution assessments at the species and family level for several macro- and meiobenthic data sets. The data of Warwick and Heip et al. were not the Resh & Unzicker case (i.e. dominance of the same higher taxa with wide pollution tolerances at both polluted and unpolluted sites), since there was essentially no difference in discrimination among sites for family- as compared to species-level analysis.

In this study we treat taxonomic sufficiency as a simple optimization problem (Pierre 1969, Walsh 1975). The performance parameter we sought to minimize was 'cost', i.e. the expertise and time needed for taxonomic identification, the variables were taxonomic levels, and the control parameter was statistical power to detect pollution impacts. The 'cost' of identification to different taxonomic levels is difficult to estimate accurately. Fortunately, the problem can be solved without knowing the actual 'costs'. Since 'cost' and level of taxonomic identification are inversely related (i.e. the higher the taxonomic level, the lower the 'cost'), the highest taxon to meet our optimization criteria for the control parameter must also be the lowest 'cost' procedure, and, therefore, the optimal solution. Optimal solutions were sought for 5 measures of community structure and 2 statistical designs employing the t -distribution. Optimal solutions, for a given sample size, took the form: "(Species, Genus, Family, Order, Phylum) is the highest taxonomic level, and, therefore, the lowest 'cost' identification procedure, necessary and sufficient to reliably ($1 - \hat{\beta} \geq 0.80$) detect significant ($\alpha = 0.05$) differences between reference and impacted conditions." Sample size requirements were evaluated in terms of statistical power and the number of pooled samples needed to obtain community-wide, unbiased estimates of 3 diversity indices.

METHODS

The data set used was from field studies conducted in 1980, 1983 and 1985 on the macrobenthos at 60 m depth along a pollution gradient in the Southern California Bight (Swartz et al. 1985, 1986, Ferraro et al. 1988). Five replicate 0.1 m² van Veen grabs were taken at each of 3 stations (Fig. 1) chosen for their ecological significance. Station R is the nearest location to the Los

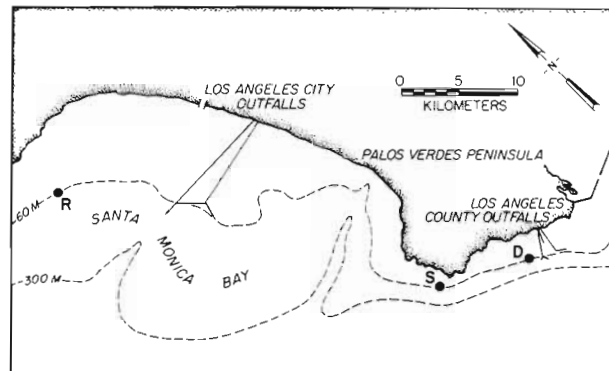


Fig. 1. Location of sampling stations in the Southern California Bight

Angeles County Sanitation District's (LACSD) outfalls representative of background conditions (Word & Mearns 1978). At Station S the macrobenthos is stimulated, having significantly more species, numerical abundance and biomass, while at Station D the macrobenthos is degraded, when compared to reference conditions (Swartz et al. 1986). Benthic community structure improved, and sediment contamination and sediment toxicity decreased along the LACSD pollution gradient in 1983 (Swartz et al. 1986), but benthic conditions retrogressed in 1985 (Ferraro et al. 1988, Stull 1988). Swartz et al. (1986) characterized benthic impacts at Stations S and D as 'major' in 1980 and 'moderate' in 1983. By comparison, benthic impacts were intermediate in 1985. In the context of the Pearson & Rosenberg (1978) model, Station S is in the 'transition zone' and Station D is near the 'peak of opportunists'.

All specimens retained on a 1.0 mm mesh sieve were identified to the lowest taxon possible, usually species. We grouped the data by taxa (species, genus, family, order, phylum), and, for each taxonomic level, we calculated 5 measures of community structure: (1) number of taxa per sample unit, (2) a Dominance Index equal to the minimum number (or fraction) of taxa whose combined abundance was equal to 75 % of the individuals in the sample unit (Swartz et al. 1985, 1986), (3) Shannon's Index, $H' = - \sum_{j=1}^s (n_j/N) \log_{10} (n_j/N)$ (Shannon & Weaver 1964)*, (4) the complement of Simpson's Index = $1 - \sum_{j=1}^s n_j (n_j - 1) / N (N - 1)$ (Simpson 1949), and (5) McIntosh's Index = $N - \sqrt{\sum_{j=1}^s n_j^2} / N - \sqrt{N}$ (McIntosh 1967), where N = total number of individuals in s groups (species, genus, family, etc.) with n_j individuals in the j th group ($j = 1, 2, \dots, s$).

* Shannon's Index was estimated by pooling samples and jackknifing (see below) Brillouin's Index, $H = 1/N \log_{10} (N! / n_1! n_2! \dots n_s!)$. See explanation for using H to estimate H' in Pielou (1966, 1975)

Means and variances for Measures 1 and 2 at each station were calculated in the standard way on untransformed and $Y^{0.5}$ -, $Y^{0.25}$ - and $\log_{10}(Y)$ -transformed data. Normality was tested by the W statistic (Shapiro & Wilk 1965), and homogeneity of variance by F_{\max} (Sokal & Rohlf 1981). We used the form of the data which best met the parametric assumption of equal variances.

Jackknife statistics (Miller 1974, Zahl 1977, Magurran 1988) were determined for Measures 3 to 5 as described in Ferraro et al. (1989). The minimum sample size (n_1 = number of replicate samples per station) for which the jackknife estimates applied was determined as the maximum value of $k + 1$ at the 3 stations, where k is the minimum number of pooled replicates whose mean index value was within $\pm 5\%$ of the index value for all (5) pooled replicate samples. The pooling procedure is a preliminary step in estimating community statistics from a sample (Pielou 1966). The asymptote criteria of $k + 1$ was chosen to ensure a stable index value and a minimum sample size of 2 for variance estimation. Our jackknife statistics are community-wide, unbiased estimates of the indices and sample size independent (for $n_1 \geq k + 1$). Jackknife statistics were computed using the computer program JACK (Rohlf 1982). Jackknife 'pseudovalues' were tested for normality by the W statistic and jackknife variances were tested for homogeneity by F_{\max} .

Our statistical model was a t -distribution for a priori paired comparisons. Statistical power was estimated for the 2-tailed (Measure 1) or the 1-tailed case (Measures 2 to 5) using the computer program POWER (Computing Center, University of Texas, Austin, TX 78712, USA). The directionality of each measure was determined by expected trends (Pearson & Rosenberg 1978, Odum 1985) and previous experience at the study site (Swartz et al. 1985, 1986, Ferraro et al. 1988). Alpha was set to 0.05. The error term for *impact* effects was the mean square among-years (MS_{ay}) at the reference station (Green 1984, 1987). The sample size for the *impact* effects tests (n_i = number of sampling years at the reference station) was varied from 5 to 2. Taxonomic sufficiency was defined for each measure as the highest taxon for which $1 - \hat{\beta} \geq 0.80$ in tests of differences between the means, or jackknife statistics, at Station R versus S and R versus D. For each contrast, we distinguished among 3 levels of response (major, intermediate, moderate) represented by the 3 sampling events (1980, 1985, 1983). Taxonomic sufficiency for *impact* effects was compared against taxonomic sufficiency for impacts inferred from differences in *location*. The error term for *location* effects was the replicate sampling error, i.e. the mean square within-stations (MS_{ws}) in ANOVA terminology. The sample size for *location* effects tests, n_l , was varied from 5 to 2. Assessing treatment effects by differences in *location* is tech-

nically inappropriate (Green 1984, 1987, Hurlbert 1984), but it is common practice in ecological field studies, and, often, there is no alternative (e.g. main sequence 4 in Green 1979).

Similarity in faunal composition was compared on the 5 taxonomic grouping levels by normal, agglomerative, hierarchical cluster analysis on untransformed and transformed ($Y^{0.5}$, $Y^{0.25}$, and $\log_{10}(Y + 1)$) abundance data using the Bray-Curtis coefficient (Bray & Curtis 1957) and the group average method (Clifford & Stephenson 1975). Results of cluster analysis were qualitatively compared with those obtained by univariate analysis on the 5 measures of community structure.

RESULTS

Our data set consisted of 60 677 individuals belonging to 14 phyla, 67 orders, 158 families, 288 genera, and 391 species (Table 1). Echinoderms were only abundant at Station R, annelids and molluscs were codominants at Station S, and annelids dominated Station D

Table 1. Number of different taxa by level of taxonomic identification^a

Phylum	Order	Family	Genus	Species
Annelida	16	42	136	197
Arthropoda	10	42	64	91
Mollusca	16	42	54	64
Rhynchocoela	4	7	8	13
Echinodermata	9	12	12	12
Cnidaria	4	4	4	4
Sipuncula	1	2	3	3
Phoronida	1	1	1	1
Echiura	1	1	1	1
Brachiopoda	1	1	1	1
Platyhelminthes	1	1	1	1
Priapulida	1	1	1	1

^a Nematoda and Bryozoa were only identified to phylum. These 2 phyla represented 0.36 % of the total numerical abundance

(Fig. 2). The numerically dominant species were *Amphiodia urtica* (% total composition = 25.8%; echinoderm) at Station R, *Tharyx* spp. (36.9%; annelid) and *Parvilucina tenuisculpta* (32.4%; mollusc) at Station S, and *Capitella* spp. (59.6%; annelid) at Station D.

Summary statistics for Measures 1 to 5 are presented in Table 2. Ninety-four percent of the means and jackknife statistics passed the normality test, while 77 % and 92 % passed the tests for homogeneity of variance within-stations and among-years, respectively. Violations of the parametric assumptions can effect power

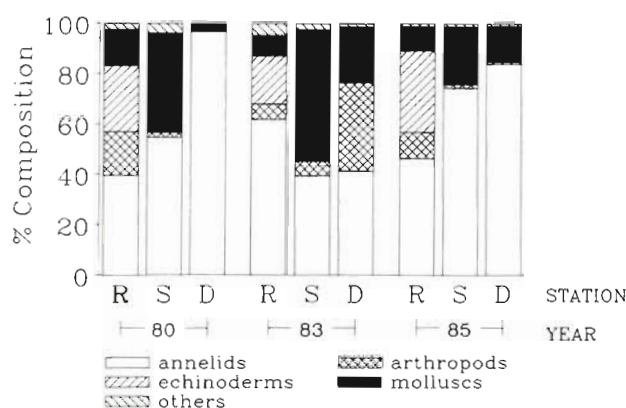


Fig. 2. Percent faunal composition by major taxa

estimates, but when, as in this study, sample sizes are equal, departures from normality are moderate, and the average among-groups variance is used to estimate high values of power (≥ 0.80), the effect is usually small (Glass et al. 1972). The effects of unusually severe winter storms and El Niño-Southern Oscillation events in southern California in 1982–83 (Dayton & Tegner 1984, Seymour et al. 1984, Stull 1988) probably made MS_{ay} a better representative estimate of natural variability than the time duration of study would suggest. The $k + 1$ values, which served as lower limits for n_i 's of the jackknife statistics, ranged from 2 to 4 (Table 3). Four replicate 0.1 m² van Veen grabs per station were needed to ensure community-wide, unbiased estimates of Shannon's, 1-Simpson's and McIntosh's Index at all of the levels of taxonomic identification investigated.

The statistical power to detect *impact* effects was dependent upon the type and magnitude of impact, and the level of taxonomic identification (Fig. 3). When $1 - \hat{\beta}$ was ≥ 0.80 for a higher taxon, it was, with few exceptions, also ≥ 0.80 for species. Reliably detectable *impact* effects based on identification to higher taxa, therefore, almost always mimicked results for species. Power was, of course, a function of n_i (not illustrated).

Number of taxa was one of the most sensitive measures of major and intermediate stimulation from the species through at least the family level of identification (Fig. 3A, C), but it was the least sensitive measure of moderate stimulation from species through order (Fig. 3B). McIntosh's and Shannon's Index were the least sensitive measures of major and intermediate stimulation, respectively, at the lower levels of taxonomic identification. All 5 measures were sensitive to major degradation, and Shannon's, 1-Simpson's and McIntosh's Index were sensitive to intermediate degradation, from species through phylum (Fig. 3D, F). Number of taxa was sensitive to moderate degradation from species to phylum, but the Dominance Index was sensitive to moderate degradation only at the species

level (Fig. 3E). The statistical power in tests of differences in *location* (Fig. 4) for a given measure and taxon was about equal to or greater than that for *impact* effects (Fig. 3) for $n_i = n_j = 5$.

Taxonomic sufficiency for *impact* effects, or the taxon with the highest power when $1 - \hat{\beta} < 0.80$ for all taxa, are listed as a function of n_i in Table 4A. All 5 measures were sensitive to major and intermediate stimulation and degradation for some taxon and n_i . Number of taxa was insensitive to moderate stimulation, while Shannon's, 1-Simpson's and McIntosh's Index were insensitive to moderate degradation, for $n_i \leq 5$. With the exception of 1-Simpson's Index, order-level identification would provide essentially the same or better detection of moderate impacts as species-level identification if one was to increase the statistical power of these insensitive measures by increasing n_i (Table 4A).

Taxonomic sufficiency for impacts inferred from differences in *location*, or the taxon with the highest power when $1 - \hat{\beta} < 0.80$ for all taxa, are listed as a function of n_i in Table 4B. As we decreased n_i , taxonomic requirements increased for some measures. Taxonomic sufficiency for number of taxa, for example, went from family ($n_i = 4$ or 5) to genus ($n_i = 3$) for Station R versus S in 1980. When taxonomic sufficiency is a function of n_i , the relative cost-efficiency of alternative sampling schemes will depend upon the incremental cost of collecting and sorting samples compared to that for taxonomically sufficient identification. The overriding criterion should be $n_i \geq k + 1$ (Table 3), however, for community-wide, unbiased estimates of the 3 diversity indices.

Cluster configurations discriminated reference, stimulated and degraded benthic conditions for the species through the order level of identification (Fig. 5). Faunal dissimilarity among station-year clusters progressively decreased for higher taxon groupings suggesting some loss of information. These results were robust when replicates were entities and for the raw data as well as the other transformations tested. Hellawell (1977), Sheehan (1984), Washington (1984), and Magurran (1988), among others, review the characteristics and performance of similarity and diversity indices in pollution studies. Taxonomic sufficiency for assessing pollution impacts based on faunal classification (Fig. 5) and the *impact* effect design (Table 4A) appeared to be similar in this study.

DISCUSSION

Our taxonomic sufficiency results apply for pollution assessments (*sensu* Edwards et al. 1975, i.e. the determination of the existence and spatial extent of an impact) at the study site using the sampling protocols

Table 2. Summary statistics [mean or jackknife statistic and mean-square within-stations (MS_{ws})] for 5 measures of community structure at Stations R, S and D in 1980, 1983 and 1985. MS_{ay} = mean square among-years at Station R

Groups: Year		Species			Genus			Number of taxa						Phyla						
		R	S	D	R	S	D	MS _{ws}	R	S	D	MS _{ws}	R	S	D	MS _{ws}				
1980	75.8	101.4	16.8	41.57 ^a	69.6	88.8	16.0	30.33 ^a	51.4	60.6	14.8	10.10	25.4	28.4 ^b	7.8	3.77	7.4	8.0	3.0	1.10
1983	71.2	73.2	35.8	176.03	65.8 ^b	67.6	33.4	155.27	51.4 ^b	48.2	27.0	80.00	27.2	22.8 ^b	15.8	17.53	8.4	6.4	5.0	1.53
1985	80.2	123.6	33.6	33.43	68.2	105.2	30.8	20.87	48.8	67.8	23.6	9.07	24.2	32.0	12.8	4.30	7.0	8.6	4.8	1.33
MS _{ay}	101.27				18.47				11.27				11.40				2.60			
Dominance Index																				
Year	R	S	D	MS _{ws}	R	S	D	MS _{ws}	R	S	D	MS _{ws}	R	S	D	MS _{ws}	R	S	D	MS _{ws}
1980	13.25	2.72	1.03 ^b	1.764 ^a	11.86	2.72	1.03 ^b	1.585 ^a	8.78	2.64	1.03 ^b	0.438 ^a	5.26	2.27	0.94 ^b	0.088	2.49	1.49	0.78	0.015 ^a
1983	15.23	3.24	7.56	10.358	13.37	3.15	7.20	10.097	9.66	2.58	5.08	7.577	5.71	1.81	3.67	1.060	1.63	1.42 ^b	1.97	0.091
1985	11.97	2.79	1.83	2.027	9.50	2.77	1.83	1.243	6.72	2.60	1.73	0.628	4.37	1.88	1.45	0.098	1.89	1.06	0.91	0.010
MS _{ay}	13.495				19.056				11.386				2.345				0.953			
Shannon's H'																				
Year	R	S	D	MS _{ws}	R	S	D	MS _{ws}	R	S	D	MS _{ws}	R	S	D	MS _{ws}	R	S	D	MS _{ws}
1980	1.491	0.940	0.350	0.0035	1.424	0.924	0.343	0.0039	1.314	0.874	0.333	0.0030	1.058	0.740	0.260	0.0019	0.619	0.415	0.068	0.0006
1983	1.607	0.902	1.217	0.0363	1.556	0.891	1.193	0.0368	1.375 ^b	0.831	1.058	0.0345	1.144 ^b	0.668	0.873	0.0166	0.525	0.440	0.492	0.0057 ^a
1985	1.357	0.937	0.634	0.0057	1.286	0.918	0.627	0.0053	1.137	0.852	0.582	0.0044	0.947	0.675	0.461	0.0024	0.543	0.298	0.223	0.0018
MS _{ay}	0.0778				0.0907				0.0765 ^a				0.0487 ^a				0.0124			
1-Simpson's Index																				
Year	R	S	D	MS _{ws}	R	S	D	MS _{ws}	R	S	D	MS _{ws}	R	S	D	MS _{ws}	R	S	D	MS _{ws}
1980	0.911	0.765	0.265	0.0016	0.906	0.764	0.265	0.0016	0.901	0.760	0.264	0.0016	0.872	0.729	0.259	0.0015	0.724	0.547	0.062	0.0007
1983	0.946	0.717	0.890	0.0088 ^a	0.942	0.718	0.888	0.0088 ^a	0.910	0.713	0.860	0.0090 ^a	0.888	0.666	0.828	0.0062 ^a	0.574 ^b	0.589	0.657	0.0088 ^a
1985	0.872	0.681	0.521	0.0038 ^a	0.868	0.681	0.521	0.0038 ^a	0.842 ^b	0.679	0.506	0.0037 ^a	0.826	0.630	0.491	0.0032 ^a	0.660	0.394	0.271	0.0031
MS _{ay}	0.0068				0.0067				0.0067				0.0053				0.0283			
McIntosh's Index																				
Year	R	S	D	MS _{ws}	R	S	D	MS _{ws}	R	S	D	MS _{ws}	R	S	D	MS _{ws}	R	S	D	MS _{ws}
1980	0.709	0.517	0.144	0.0008	0.700	0.516	0.144	0.0008	0.692	0.512	0.144	0.0008	0.649	0.481	0.140	0.0007	0.479	0.328	0.032	0.0004
1983	0.772	0.467	0.676	0.0098	0.763	0.467	0.674	0.0095	0.706	0.463	0.634	0.0093	0.671	0.422	0.593	0.0056	0.349 ^b	0.360	0.420	0.0057 ^a
1985	0.649	0.437	0.309	0.0023	0.643	0.437	0.309	0.0023	0.609	0.435	0.299	0.0021	0.587	0.393	0.288	0.0017	0.421	0.222	0.147	0.0013
MS _{ay}	0.0190				0.0182				0.0136				0.0095				0.0213			

^a Probability of homogeneous variances <0.01; ^b probability of normal distribution <0.01

^a Probability of homogeneous variances < 0.01; ^b probability of normal distribution < 0.01

Table 3. Minimum number of pooled 0.1 m² van Veen grab samples needed to meet our asymptote criterion of $k + 1$ samples for 3 diversity indices

Taxon	Year	Shannon's H'	McIntosh's Index	1-Simpson's Index
Species	1980	4	4	3
	1983	4	2	3
	1985	3	3	2
Genus	1980	3	4	3
	1983	4	2	3
	1985	3	3	2
Family	1980	3	4	3
	1983	4	2	3
	1985	3	3	2
Order	1980	3	3	3
	1983	3	3	3
	1985	2	3	2
Phylum	1980	2	4	3
	1983	3	4	4
	1985	2	3	3

tested. One should not assume they apply for other objectives, experimental and statistical designs, or sampling protocols at the study site or at some other location. There are also many other measures of community structure (e.g. Washington 1984), some of which may be more sensitive than those we tested. Our findings have general significance in that they support accumulating evidence (e.g. Heip et al. 1988, Warwick 1988a, b, Kingston & Riddle 1989) indicating that taxonomic sufficiency is a useful concept in pollution ecology.

An appropriate statistical design is an important element of any quantitative study. Some problems of environmental field experiments and impact assessments (see Hurlbert 1984) are (arguably) addressed by a variety of statistical designs (Moore & McLaughlin 1978, Green 1979, 1984, 1987, Skalski & McKenzie 1982, Bernstein & Zalinski 1983, Mar et al. 1985, Millard & Lettenmaier 1986, Stewart-Oaten et al. 1986, Walters et al. 1988). The more exact designs are difficult to implement, having formidable data requirements, such as long-term pre-impact and post-impact (BACI design; Stewart-Oaten et al. 1986) or frequent interval temporal ('staircase' design; Walters et al. 1988) observations, which are rarely met. Since data constraints often compel the use of suboptimal designs, it is important to know if they provide adequate approximate solutions.

We compared the statistical power and taxonomic sufficiency for pollution assessments using 2 statistical designs: a *t*-test for differences between reference and impacted stations where the error term was the (1)

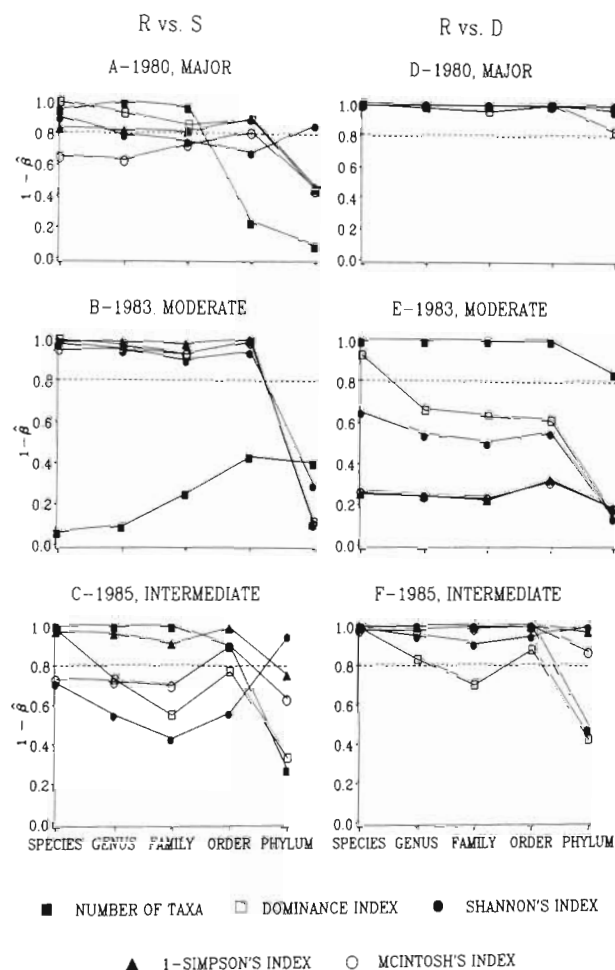


Fig. 3. Estimated statistical power ($1 - \hat{\beta}$) as a function of the level of taxonomic identification in tests of impact effects for Station R versus S (A to C), and Station R versus D (D to F). Alpha = 0.05, and the number of sampling years at the reference station, $n_i = 5$

among-year variation at the reference station (MS_{ay}) (impact effects design), and (2) replicate sampling error (MS_{ws}) (location effects design). MS_{ws} is an estimate of the small-scale (meters to a few decameters) spatial variability in our community structure measures at our 3 stations. The temporal variability at our reference station, MS_{ay} , is a more meaningful measure of the natural variability, and, therefore, the more appropriate error term for detecting deviations from natural conditions (Green 1984, 1987). The duration and frequency of sampling required to determine the normal range of biological variability of an unperturbed system will vary (Green 1979). Even if sampling encompasses the natural range of variability, the ability to statistically discriminate impacted from reference conditions still depends upon the magnitude of impact one wishes to detect and the sample size (n_i , in the impact effects design). Our impact effect results (Table 4A) suggest

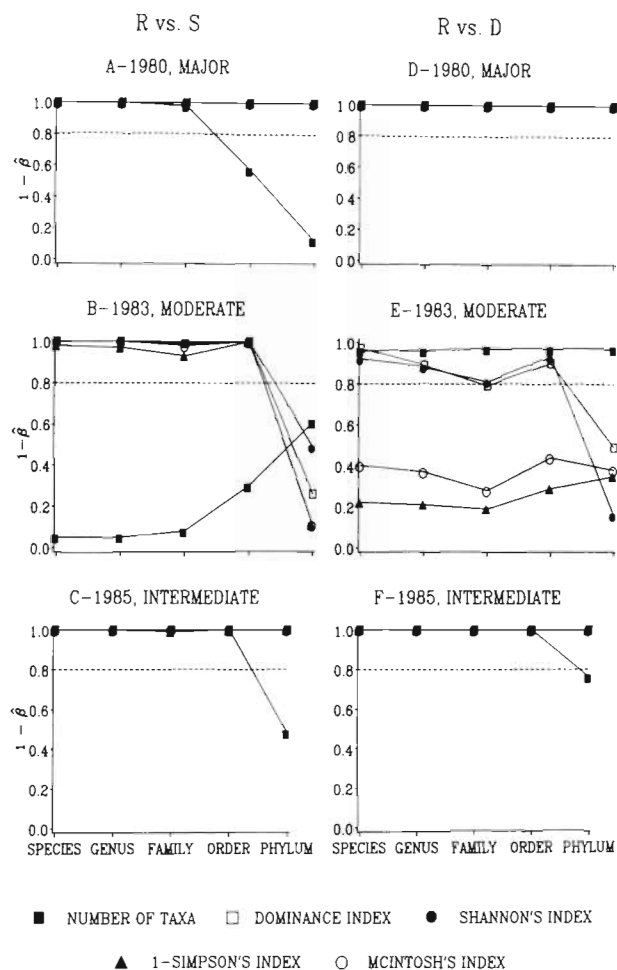


Fig. 4. Estimated statistical power ($1 - \hat{\beta}$) as a function of the level of taxonomic identification in tests of *location* effects for Station R versus S (A to C), and Station R versus D (D to F). Alpha = 0.05, and the number of replicate samples per station, $n_1 = 5$

that 4 yr sampling at the reference station and family-level identification is sufficient for detecting intermediate or larger impacts on number of taxa and several other measures of community structure. Other measures or more years of annual sampling may be needed to reliably detect lesser impacts. According to Gray (1981), decades of annual sampling may be needed to establish a baseline against which to assess the chronic effects of low levels of pollutants on benthic communities.

In this study MS_{ay} was usually greater than MS_{ws} (Table 2), and the statistical power of measures capable of reliably detecting differences by *location* effects was almost always greater than that by *impact* effects when $n_1 = n_i$ (Figs. 3 & 4). The only times MS_{ws} substantially exceeded MS_{ay} was for number of taxa in 1983, the year of unusual and severe climatic events and moderate benthic impacts. When variances among stations

are homogeneous, the expectation of MS_{ay} ($E(MS_{ay})$) will always be $\geq E(MS_{ws})$, since MS_{ay} subsumes MS_{ws} , because, in practice, one never samples in exactly the same spot each year. Benthic impact assessments using the *location* effects design, therefore, will generally be liberal (i.e. more environmentally protective) than assessments using the more appropriate *impact* effects design (when $n_1 = n_i$). We recommend that impact determinations based on *location* effects be considered tentative until confirmed by replicate temporal sampling and implementation of the *impact* effects design, or, if possible, one of the more exact designs cited above. We do not recommend using the among-station variability (MS_{as}) of presumed unimpacted stations over a wide geographic area as the error term in tests of pollution impacts (e.g. Scanland 1987). MS_{as} is a measure of biogeographic variability (and variability due to habitat heterogeneity if samples are not taken from identical habitats) rather than the natural variability of the parameters of interest at the study site.

The statistical explanation for taxonomic sufficiency using the *impact* and *location* effects designs is

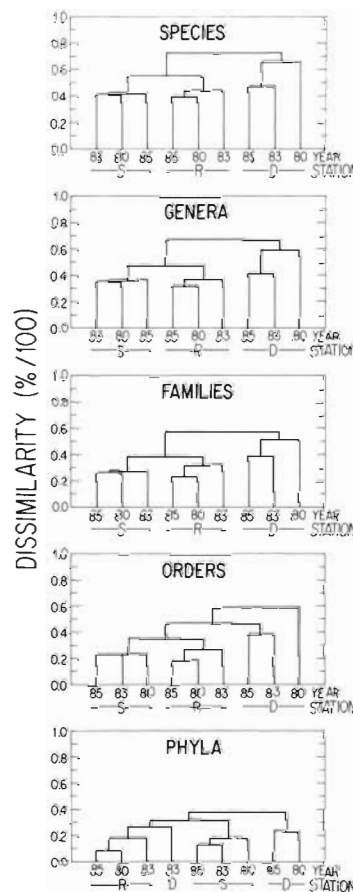


Fig. 5. Dendrograms of normal, agglomerative, hierarchical cluster analysis on $Y^{0.25}$ -transformed abundance data of pooled replicate samples using the Bray-Curtis coefficient and the group average method

Table 4A & B. In capital letters, the highest taxonomic level needed to reliably ($1-\beta \geq 0.80$) detect a significant ($\alpha = 0.05$) difference between reference and stimulated (R vs S) or degraded (R vs D) macrobenthic communities at 60 m depth in the Southern California Bight (USA). In lower case letters, the taxon yielding the highest power when $1-\beta < 0.80$ for all taxa. Impacts were characterized as major, moderate and intermediate in 1980, 1983 and 1985, respectively. Samples were 0.1 m² van Veen grabs sieved through a 1.0 mm mesh screen. A. *Impact* effects were determined by a comparison of means, or jackknife statistics, by *t*-tests, where the error term was the mean square among-years at the reference station, and sample size = n_i = number of sampling years at the reference station. B. *Location* effects were determined by a comparison of means, or jackknife statistics, by *t*-tests, where the error term was the mean square within-stations, and sample size = n_i = number of replicate samples per station

Measure		R vs S			R vs D		
		Major 1980	Moderate 1983	Intermed. 1985	Major 1980	Moderate 1983	Intermed. 1985
A. Impact effects							
Number taxa	n_i						
	5	FAMILY	order	ORDER	PHYLUM	PHYLUM	ORDER
	4	FAMILY	order	FAMILY	PHYLUM	ORDER	ORDER
	3	GENUS	order	FAMILY	ORDER	ORDER	ORDER
Dominance Index	2	genus	order	FAMILY	FAMILY	FAMILY	FAMILY
	5	ORDER	ORDER	SPECIES	PHYLUM	SPECIES	ORDER
	4	ORDER	ORDER	SPECIES	ORDER	SPECIES	SPECIES
	3	SPECIES	ORDER	SPECIES	ORDER	species	SPECIES
Shannon's Index (H')	2	species	species	species	species	species	species
	5	PHYLUM	ORDER	PHYLUM	PHYLUM	species	PHYLUM
	4	SPECIES	ORDER	PHYLUM	PHYLUM	species	PHYLUM
	3	species	species	phylum	PHYLUM	species	PHYLUM
1-Simpson's Index	2	species	species	phylum	PHYLUM	species	phylum
	5	ORDER	ORDER	ORDER	PHYLUM	order	PHYLUM
	4	ORDER	ORDER	ORDER	PHYLUM	order	PHYLUM
	3	order	ORDER	ORDER	PHYLUM	order	ORDER
McIntosh's Index	2	order	order	order	PHYLUM	order	ORDER
	5	ORDER	ORDER	ORDER	PHYLUM	order	PHYLUM
	4	order	ORDER	ORDER	PHYLUM	order	ORDER
	3	order	ORDER	order	PHYLUM	order	ORDER
	2	order	order	order	ORDER	order	order
	B. Location effects						
	n_i						
	Number taxa	5	FAMILY	phylum	ORDER	PHYLUM	PHYLUM
4		FAMILY	phylum	ORDER	PHYLUM	PHYLUM	ORDER
3		GENUS	phylum	ORDER	PHYLUM	phylum	ORDER
2		species	phylum	FAMILY	ORDER	phylum	FAMILY
Dominance Index	5	PHYLUM	ORDER	PHYLUM	PHYLUM	ORDER	PHYLUM
	4	PHYLUM	ORDER	PHYLUM	PHYLUM	ORDER	PHYLUM
	3	PHYLUM	ORDER	PHYLUM	PHYLUM	species	PHYLUM
	2	PHYLUM	order	PHYLUM	PHYLUM	species	PHYLUM
Shannon's Index (H')	5	PHYLUM	ORDER	PHYLUM	PHYLUM	ORDER	PHYLUM
	4	PHYLUM	ORDER	PHYLUM	PHYLUM	ORDER	PHYLUM
	3	PHYLUM	ORDER	PHYLUM	PHYLUM	order	PHYLUM
	2	PHYLUM	order ^a	PHYLUM	PHYLUM	order ^a	PHYLUM
1-Simpson's Index	5	PHYLUM	ORDER	PHYLUM	PHYLUM	phylum	PHYLUM
	4	PHYLUM	ORDER	PHYLUM	PHYLUM	phylum	PHYLUM
	3	PHYLUM	ORDER	PHYLUM	PHYLUM	phylum ^a	PHYLUM
	2	PHYLUM ^a	order ^a	PHYLUM ^a	PHYLUM ^a	phylum ^a	PHYLUM ^a
McIntosh's Index	5	PHYLUM	ORDER	PHYLUM	PHYLUM	order	PHYLUM
	4	PHYLUM	ORDER	PHYLUM	PHYLUM	order	PHYLUM
	3	PHYLUM ^a	ORDER	PHYLUM	PHYLUM ^a	order	PHYLUM
	2	PHYLUM ^a	order ^a	PHYLUM ^a	PHYLUM ^a	order ^a	PHYLUM ^a
^a These sample sizes are not recommended since $n_i < k + 1$ (Table 3); $n_i \geq 4$ to ensure community-wide, unbiased estimates of the diversity indices at all taxonomic levels investigated							

^a These sample sizes are not recommended since $n_i < k + 1$ (Table 3); $n_i \geq 4$ to ensure community-wide, unbiased estimates of the diversity indices at all taxonomic levels investigated

Table 5. Critical 'effect size' (ES) values for 1- and 2-tailed tests when $\alpha = 0.05$, $1-\beta = 0.80$, and sample size = 2 to 5

n	1-tail	2-tail
5	1.69	2.00
4	1.96	2.36
3	2.45	3.05
2	3.95	5.63

straightforward. A given taxonomic level is sufficient to address a particular objective when $ES \geq$ a critical value, determined by the power criterion for taxonomic sufficiency (in our example $1-\beta \geq 0.80$), α , and n_i or n_j (Table 5). As long as the mean difference between the reference and impacted station (numerator of ES) and the SD (denominator of ES) remain approximately proportional (i.e. $ES \cong \text{constant}$) for the different taxonomic levels, the statistical power of the test will vary little.

The biological explanation (with some caveats) for taxonomic sufficiency in pollution impact assessments, we believe, is the hierarchical structure of biological responses to stress (Pearson & Rosenberg 1978, Boesch & Rosenberg 1981). As stress increases, the adaptability of first the individual, then the species, genus, family, etc. is exceeded. Consequently, impacts resulting from increasing stress are manifest at higher and higher levels of biological organization, and the taxonomic level sufficient to assess impacts should increase in a step function for a given sample size. Our results for degradation effects (Fig. 3; Table 4A) show this relationship. A good example is the Dominance Index, where taxonomic sufficiency was phylum, order and species for major, intermediate and moderate degradation, respectively, for $n_i = 5$ (Table 4A). The step relationship, however, did not hold for stimulation effects, probably due to the nonlinear response of several of our community structure measures to organic enrichment (Pearson & Rosenberg 1978, Gray 1981). Nevertheless, higher taxa were often sufficient for assessing enrichment effects in our study (Table 4).

According to the hierarchic-response-to-stress theory, species should be the most sensitive taxon for assessing pollution impacts. However changes in species populations in response to a small pollutant stress may be difficult or impossible to detect using standard sampling protocols due to the large natural variability in the numerical abundances of even common macrobenthic species (Eagle 1975, LACSD 1981, Nichols 1985, Stull et al. 1986). Grouping animals to genus (or higher taxa) may dampen the natural variability in the number of taxa and other measures of community structure (Table 2), thus actually improving our ability to assess small pollution impacts.

There is also the practical matter of the precision and accuracy of taxonomic identification (Ellis & Cross 1981, Ellis 1985). Species identification is more error prone than identification to higher taxa (Ellis & Cross 1981), and correct identification to a higher taxon may be preferable, both biologically and statistically, than incorrect identification to species (Green 1979).

Taxonomic sufficiency is a pragmatic construct in which the accuracy of identification is balanced against the need for information (Ellis 1985). Information obtained by more accurate taxonomic identification is considered redundant when the same inference is drawn when animals are identified to a higher taxon (Warwick 1988b). Taxonomic sufficiency depends upon the objective, the study site, the measure, the statistical design, and the sample size (Table 4). It may also depend upon other sampling variables (e.g. sample unit size, sieve mesh size) which were kept constant in this study. Taxonomic demands are important to consider when designing a cost-effective sampling program since laboratory sample-processing costs are primarily responsible for the difference in cost of alternative macrobenthic sampling schemes (Saila et al. 1976, Kingston & Riddle 1989). About 60 % of the total laboratory processing time in this study was for species identification and counting of animals (Ferraro et al. 1988). This percentage will vary depending on the study and laboratory personnel (Ferraro et al. 1989). But, clearly, taxonomic identification is a major cost factor in many macrobenthic studies, and the time and cost of identification can be greatly reduced if animals are identified only to genus, family, order or phylum (McIntyre et al. 1984, Warwick 1988b).

Information such as in Figs. 3 & 4 and Tables 3 & 4 can be used to determine the reasonableness of a study's objective(s) and the components of an optimal experimental design. The individual investigator must decide which and how many response measures should be required to meet a particular objective. In this study, an $n_i \geq 4$ was needed to ensure a high probability of detecting intermediate or larger impacts at the family or higher level of identification on all 5 measures of community structure using the *location* effects design (Table 4B). Four replicate 0.1 m² van Veen grab samples per station were needed to ensure obtaining community-wide estimates of 3 diversity indices at all the taxonomic levels investigated (Table 3). Several replicate samples per station are also likely to provide better representative estimates of the number of taxa and the Dominance Index when employing the *impact* effects design. We, therefore, recommend that an $n_i \geq 4$ be used when assessing impacts at the study site using both the *location* and *impact* effects design. Although order and $n_i \geq 3$ appears sufficient for many purposes, in our opinion, the *impact* effects design with $n_i \geq 4$ and

family-level identification would be a good choice for assessing pollution impacts at our study site. This design would ensure a high probability ($1 - \beta \geq 0.80$) of detecting major to intermediate stimulation and degradation effects on the number of taxa (probably the most widely accepted measure of community structure) and several other measures of community structure (Table 4A). Furthermore, taxonomic keys for families of major taxa are readily available, and family-level identification is relatively easy for a non-specialist to master (McIntyre et al. 1984, Fauchald 1989).

Warwick's (1988a, b) contention that analysis of pollution effects on higher taxa minimizes confounding effects of so-called 'nuisance' variables (e.g. water depth and sediment granulometry) relies on the assumption that 'these [nuisance] variables usually influence the fauna more by species replacement than by changes in the proportions of major taxa present'. This assumption needs to be tested. The hierarchic-response-to-stress theory applies to natural and pollution stressors, and the relative influence of multiple stressors on benthic communities is likely to depend upon their relative strengths. Clearly, water depth and grain size *can* have as great or greater effect on benthic community structure as pollution (Smith & Greene 1976, Ward & Young 1982, Gray et al. 1988). Field experiments using artificial substrates (Cairns & Pratt 1986) and laboratory mesocosm experiments (Gray et al. 1988) are probably the best ways to address the potential confounding of pollution and environmental factors.

There are pollution research objectives for which species-level identification may be either required or very desirable. If an impact has occurred, one is often interested in determining the specific cause(s) (e.g. Chemical A) and understanding the mode of action on individuals or species (e.g. toxicity, impairment of growth or reproduction, life history adaptations). A species list contains much implicit information about the ecosystem state (Botkin et al. 1979) which might be used to infer causality given sufficient knowledge of species niches, including chemical tolerances. The same principle holds for higher taxa listings, but as the taxonomic level increases the niche space broadens, and the resolution decreases. Developing ecologic theory, and some empirical evidence, indicates that information at lower levels of biological organization (individual, species, metasppecies) is necessary for understanding ecosystem responses to anthropogenic stress (Grassle & Grassle 1974, 1977, Schindler 1987, Underwood & Peterson 1988).

In conclusion, taxonomic sufficiency appears to be a useful concept in pollution ecology. The level of taxonomic identification and sample size should be considered along with other sampling variables (e.g. sample

unit size, sieve mesh size) when seeking a statistically rigorous, cost-effective sampling design sufficient to meet pollution assessment objectives. As empirical information on the taxonomic level needed for pollution assessments accumulates, it may be possible to accurately predict the taxonomic sufficiency for a given habitat, pollution condition, and sampling scheme.

Acknowledgements. We thank Rick Swartz for use of his 1980 and 1983 data, Rick Swartz, Jim Heltshe and Ted DeWitt for their reviews of the manuscript, and Jim Bertling for computer support. Contribution No. N-124 of the US EPA, Environmental Research Laboratory, Narragansett, RI, USA.

LITERATURE CITED

- Bernstein, B. B., Zalinski, J. (1983). An optimum sampling design and power tests for environmental biologists. *J. environ. Manage.* 16: 35–43
- Boesch, D. F., Rosenberg, R. (1981). Response to stress in marine benthic communities. In: Barrett, G. W., Rosenberg, R. (eds.) *Stress effects on natural ecosystems*. Wiley, New York, p. 179–200
- Botkin, D. B., Maguire, B., Moore, B., III, Morowitz, H. J., Slobodkin, L. B. (1979). A foundation for ecological theory. *Memorie Ist. ital. Idrobiol.*, 37 (Suppl.): 13–31
- Bray, J. R., Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* 27: 325–349
- Cairns, J., Jr, Pratt, J. R. (1986). Developing a sampling strategy. In: Isom, B. G. (ed.) *Rationale for sampling and interpretation of ecological data in the assessment of freshwater ecosystems*. ASTM STP 894, American Testing Society for Testing and Materials, Philadelphia, p. 168–186
- Clifford, H. T., Stephenson, W. (1975). An introduction to numerical classification. Academic Press, New York
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*, revised edn. Academic Press, Orlando, Florida
- Dayton, P. K., Tegner, M. J. (1984). Catastrophic storms, El Niño, and patch stability in a southern California kelp community. *Science* 224: 283–285
- Eagle, R. A. (1975). Natural fluctuations in a soft bottom benthic community. *J. mar. biol. Ass. U.K.* 55: 865–878
- Edwards, R. W., Hughes, B. D., Read, M. W. (1975). Biological survey in the detection and assessment of pollution. In: Chadwick, M. J., Goodman, G. T. (eds.) *The ecology of resource degradation and renewal. The 15th Symposium of the British Ecological Society*, 10–12 July 1973. Wiley, New York, p. 139–156
- Ellis, D. (1985). Taxonomic sufficiency in pollution assessment. *Mar. Pollut. Bull.* 16: 459
- Ellis, D. V., Cross, S. F. (1981). A protocol for inter-laboratory calibrations of biological species identifications (ring tests). *Wat. Res.* 15: 1107–1108
- Fauchald, K. (1989). The Second Annual Riser Lecture: eclecticism and the study of polychaetes. *Proc. Biol. Soc. Wash.* 102: 742–752
- Ferraro, S. P., Swartz, R. C., Cole, F. A., DeBen, W. A. (1988). Quantitative macrobenthic sampling: sample unit size and number, statistical power and 'cost'. Final Report to the Office of Marine and Estuarine Protection, U.S. Environmental Protection Agency, 401 M Street, S.W. Washington, D.C. 20460

- Ferraro, S. P., Cole, F. A., DeBen, W. A., Swartz, R. C. (1989). Power-cost efficiency of eight macrobenthic sampling schemes in Puget Sound, Washington, USA. *Can. J. Fish. Aquat. Sci.* 46: 2157–2165
- Glass, G. V., Peckham, P. D., Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *R. Educ. Res.* 42: 237–288
- Grassle, J. F., Grassle, J. P. (1974). Opportunistic life histories and genetic systems in marine benthic polychaetes. *J. mar. Res.* 32: 253–284
- Grassle, J. F., Grassle, J. P. (1977). Temporal adaptations in sibling species of *Capitella*. In: Coull, B. C. (ed.) *Ecology of marine benthos*. Belle W. Baruch Library in Marine Science No. 6, University of South Carolina Press, Columbia, p. 177–189
- Gray, J. S. (1981). *The ecology of marine sediments. An introduction to the structure and function of benthic communities*. Cambridge University Press, Cambridge
- Gray, J. S., Aschan, M., Carr, M. R., Clarke, K. R., Green, R. H., Pearson, T. H., Rosenberg, R., Warwick, R. M. (1988). Analysis of community attributes of the benthic macrofauna of Frierfjord/Langesundfjord and in a mesocosm experiment. *Mar. Ecol. Prog. Ser.* 46: 151–165
- Green, R. H. (1979). *Sampling design and statistical methods for environmental biologists*. Wiley, New York
- Green, R. H. (1984). Statistical and nonstatistical considerations for environmental monitoring studies. *Envir. Monitg Assessmt* 4: 293–301
- Green, R. H. (1987). Statistical and mathematical aspects: distinction between natural and induced variation. In: Vouk, V. B., Butler, G. C., Upton, A. C., Parke, D. V., Asher, S. C. (eds.) *Methods for assessing the effects of mixtures of chemicals*. Wiley, Chichester, p. 335–354
- Heip, C., Warwick, R. M., Carr, M. R., Herman, P. M. J., Huys, R., Smol, N., van Holsbeke, K. (1988). Analysis of community attributes of the benthic meiofauna of Frierfjord/Langesundfjord. *Mar. Ecol. Prog. Ser.* 46: 171–180
- Hellawell, J. M. (1977). Change in natural and managed ecosystems: detection, measurement and assessment. *Proc. R. Soc. Lond. B* 197: 31–57
- Hughes, B. D. (1978). The influence of factors other than pollution on the value of Shannon's diversity index for benthic macro-invertebrates in streams. *Wat. Res.* 12: 359–364
- Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* 54: 187–211
- Kingston, P. F., Riddle, M. J. (1989). Cost effectiveness of benthic faunal monitoring. *Mar. Pollut. Bull.* 20: 490–496
- Los Angeles County Sanitation Districts (LACSD) (1981). *Ocean monitoring and research. Annual Report 1980–1981. Chapter III. Benthic physical, chemical and biological conditions*. Los Angeles County Sanitation Districts, Whittier, California, p. 22–150
- Magurran, A. E. (1988). *Ecological diversity and its measurement*. Princeton University Press, Princeton, New Jersey
- Mar, B. W., Lettenmaier, D. P., Horner, R. R., Richey, J. S., Palmer, R. N., Millard, S. P., MacKenzie, M. C., Vega-Gonzalez, S., Lund, J. R. (1985). *Sampling design for aquatic ecological monitoring. Vol. 1: Summary Report*. Electric Power Research Institute Project 1729-1, Department of Civil Engineering, University of Washington, Seattle, WA. (Available from Research Reports Center, Box 50490, Palo Alto, CA 94303)
- McIntosh, R. P. (1967). An index of diversity and the relation of certain concepts to diversity. *Ecology* 48: 392–404
- McIntyre, A. D., Elliott, J. M., Ellis, D. V. (1984). Introduction: design of sampling programmes. In: Holme, N. A., McIntyre, A. D. (eds.) *Methods for the study of marine benthos*. Blackwell Scientific Publications, Oxford, p. 1–26
- Millard, S. P., Lettenmaier, D. P. (1986). Optimal design of biological sampling programs using the analysis of variance. *Estuar. coast. Shelf Sci.* 22: 637–656
- Miller, R. G. (1974). The jackknife – a review. *Biometrika* 61: 1–15
- Moore, S. F., McLaughlin, D. B. (1978). Design of field experiments to determine the ecological effects of petroleum in intertidal ecosystems. *Wat. Res.* 12: 1091–1099
- Nichols, F. H. (1985). Abundance fluctuations among benthic invertebrates in two Pacific estuaries. *Estuaries* 8: 136–144
- Odum, E. P. (1985). Trends expected in stressed ecosystems. *BioScience* 35: 419–422
- Pearson, T. H., Rosenberg, R. (1978). Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. *Oceanogr. mar. Biol. A. Rev.* 16: 229–311
- Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *J. Theor. Biol.* 13: 131–144
- Pielou, E. C. (1975). *Ecological diversity*. Wiley, New York
- Pierre, D. A. (1969). *Optimization theory with applications*. Wiley, New York
- Resh, V. H., Unzicker, J. D. (1975). Water quality monitoring and aquatic organisms: the importance of species identification. *J. Wat. Pollut. Control Fed.* 47: 9–19
- Rohlf, F. J. (1982). *BIOM a package of statistical programs*. Department of Ecology and Evolution, SUNY at Stony Brook, New York 11794
- Saila, S. B., Pikanowski, R. A., Vaughan, D. S. (1976). Optimum allocation strategies for sampling benthos in the New York Bight. *Estuar. coast. mar. Sci.* 4: 119–128
- Scanland, T. B. (1987). A comparison of the infauna of the Palos Verdes with the SCCWRP 1985 Control Survey results. In: *Joint Water Pollution Control Plant 1988 Revised Application for Modification of Secondary Treatment Requirements for Discharges into Marine Waters. Vol. II, January 20, 1988, Appendix F-9, 11 p. + 1 table & 3 figs.* (Available from Los Angeles County Sanitation Districts, Whittier, CA 90607)
- Schindler, D. W. (1987). Detecting ecosystem responses to anthropogenic stress. *Can. J. Fish. Aquat. Sci.* 44 (Suppl. 1): 6–25
- Seymour, R. J., Strange, R. R. III, Cayan, D. R., Nathan, R. A. (1984). Influence of El Niños on California's wave climate. In: Edge, B. L. (ed.) *Nineteenth Coastal Engineering Conference, Proceedings of the International Conference, September 3–7, 1984, Houston, Texas*. American Society of Civil Engineers, New York, I: 577–592
- Shannon, C. E., Weaver, W. (1964). *The mathematical theory of communication*. University of Illinois Press, Urbana, Illinois
- Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* 52: 591–611
- Sheehan, P. J. (1984). Effects on community and ecosystem structure and dynamics. In: Sheehan, P. J., Miller, D. R., Butler, G. C., Bourdeau, Ph. (eds.) *Effects of pollutants at the ecosystem level*. Wiley, Chichester, p. 51–99
- Simpson, E. H. (1949). Measurement of diversity. *Nature, Lond.* 163: 688
- Skalski, J. R., McKenzie, D. H. (1982). A design for aquatic monitoring programs. *J. environ. Mgmt* 14: 237–251
- Smith, R. W., Greene, C. S. (1976). Biological communities near submarine outfall. *J. Wat. Pollut. Control Fed.* 48: 1894–1912
- Sokal, R. R., Rohlf, F. J. (1981). *Biometry. The principles and*

- practice of statistics in biological research, 2nd edn. W. H. Freeman and Company, San Francisco
- Stewart-Oaten, A., Murdoch, W. W., Parker, K. R. (1986). Environmental impact assessment: 'pseudoreplication' in time? *Ecology* 67: 929–940
- Stull, J. K. (1988). Discussion of benthic biological long-term trends results. In: Joint Water Pollution Control Plant 1988 Revised Application for Modification of Secondary Treatment Requirements for Discharges into Marine Waters. Vol. II, January 20, 1988, Appendix F-4, 16 p. + 18 figs. (Available from Los Angeles County Sanitation Districts, Whittier, CA 90607)
- Stull, J. K., Haydock, C. I., Montagne, D. E. (1986). Effects of *Listriolobus pelodes* (Echiura) on coastal shelf benthic communities and sediments modified by a major California wastewater discharge. *Estuar. coast. Shelf Sci.* 22: 1–17
- Swartz, R. C., Cole, F. A., Schults, D. W., DeBen, W. A. (1986). Ecological changes in the Southern California Bight near a large sewage outfall: benthic conditions in 1980 and 1983. *Mar. Ecol. Prog. Ser.* 31: 1–13
- Swartz, R. C., Schults, D. W., Ditsworth, G. R., DeBen, W. A., Cole, F. A. (1985). Sediment toxicity, contamination, and macrobenthic communities near a large sewage outfall. In: Boyle, T. P. (ed.) Validation and predictability of laboratory methods for assessing the fate and effects of contaminants in aquatic ecosystems. ASTM STP 865, American Testing Society for Testing and Materials, Philadelphia, p. 152–175
- Underwood, A. J., Peterson, C. H. (1988). Towards an ecological framework for investigating pollution. *Mar. Ecol. Prog. Ser.* 46: 227–234
- Walsh, G. R. (1975). *Methods of optimization*. Wiley, London
- Walters, C. J., Collie, J. S., Webb, T. (1988). Experimental designs for estimating transient responses to management disturbances. *Can. J. Fish. Aquat. Sci.* 45: 530–538
- Ward, T. J., Young, P. C. (1982). Effects of sediment trace metals and particle size on the community structure of epibenthic seagrass fauna near a lead smelter, South Australia. *Mar. Ecol. Prog. Ser.* 9: 137–146
- Warwick, R. M. (1988a). The level of taxonomic discrimination required to detect pollution effects on marine benthic communities. *Mar. Pollut. Bull.* 19: 259–268
- Warwick, R. M. (1988b). Analysis of community attributes of the macrobenthos of Frierfjord/Langesundfjord at taxonomic levels higher than species. *Mar. Ecol. Prog. Ser.* 46: 167–170
- Washington, H. G. (1984). Diversity, biotic and similarity indices. A review with special relevance to aquatic ecosystems. *Wat. Res.* 18: 653–694
- Word, J. Q., Mearns, A. J. (1978). The 60-meter control survey. Annual Report 1978, Southern California Coastal Water Research Project, El Segundo, California, p. 41–56
- Wu, R. S. S. (1982). Effects of taxonomic uncertainty on species diversity indices. *Mar. environ. Res.* 6: 215–225
- Zahl, S. (1977). Jackknifing an index of diversity. *Ecology* 58: 907–913

This article was presented by Professor J. Gray, Oslo, Norway

Manuscript first received: March 6, 1990

Revised version accepted: July 17, 1990