Application of the Great Lakes National Program Office's Data Quality Objective to Benthos Data Generated by the Annual Water Quality Survey

Richard P. Barbiero

CSC
1359 West Elmdale Avenue
Suite #2
Chicago, Illinois 60660

Prepared for:
United States Environmental Protection Agency
Great Lakes National Program Office
77 West Jackson Boulevard
Chicago, Illinois 60604

Louis Blume, Project Officer

February 2003

Acknowledgements

This report was prepared under the direction of Louis Blume, Project Officer, Great Lakes National Program Office (EPA Contract No. 68-C-01-091).

Summary

A Data Quality Objective (DQO) has been developed by the Great Lakes National Program Office (GLNPO) to ensure that data collected from their Water Quality Surveys are of suitable quality to provide decision makers with sufficient certainty to make educated ecological management decisions. The current GLNPO DQO states that data quality should be sufficient for there to be an 80% chance of detecting a 20% change, at the 90% confidence level, between current and historical measurements of a variable made in a particular lake during a particular season.

This report assesses the extent to which benthos data collected in summer, 1999 comply with the GLNPO DQO. The most important findings are summarized below:

- Sufficient inter-site variation exists between offshore stations within each lake basin to consider each station representative of a separate statistical population, rather than a replicate of a larger, basin-wide statistical population.
- When densities of the most dominant species/taxonomic groups were examined, the target contained in the DQO was by and large not met by the present sampling effort.
- Sample sizes required to meet the DQO when just the most dominant species/taxonomic groups are considered are unfeasibly large (> 12 in almost all cases; > 35 in the majority of cases).
- Variation was in some cases higher at nearshore than offshore stations. However, because of the higher overall densities of organisms at nearshore stations, sample sizes required to satisfy the DQO tended to be slightly higher at deeper stations.
- The current detection target of a 20% change is much lower than actual interannual differences seen in *Diporeia* densities between 1997 and 2001, even where no consistent trends (e.g., declines) were noted. Changes this small are therefore likely to be within the range of natural fluctuation for most benthic organism populations, and as such are probably of limited inherent ecological interest or use.
- In spite of the inability to meet the DQO, the current level of replication was sufficient to detect interannual changes in *Diporeia* densities from 1997 to 2001 at almost all sites for which data were available.
- In general, the current DQO is ambiguous with regard to specifically which differences are of interest. Several interpretations are possible, each requiring a different statistical approach.
- The current DQO does not address which biological variables are of interest, and therefore which should be subject to its specifications.

1 Introduction

1.1 Benthos program

The Great Lakes National Program Office of the US EPA has been involved in regular surveillance monitoring of the open waters of the Laurentian Great Lakes since 1983. This surveillance monitoring is meant to satisfy the provisions of the Great Lakes Water Quality Agreement (International Joint Commission 1978), which calls for periodic monitoring of the lakes to evaluate water quality trends over time. In 1997, a benthic invertebrate monitoring program was added to complement GLNPO's existing open water program. This program differs from the open water surveys in a number of important ways. Given the homogeneity of most of the open waters of the Great Lakes, the open water survey is designed to detect changes on a fairly large spatial scale. The statistical populations which the sampling program of this survey were designed to estimate correspond to lake basins, thus the half dozen or so sampling stations established within each basin serve as replicates. In contrast, the benthic program originally employed a sampling strategy designed to characterize communities from two habitat types, the nearshore (<50 m) and offshore (>50 m). The rationale behind this design was that the offshore benthic communities would serve as integrators of conditions on a larger, basin-wide scale, while nearshore locations would exhibit a stronger dependence on local conditions and offer a better indication of relatively short-term responses to local stressors. While representative stations were established in both offshore and nearshore locations, coverage was not sufficient to provide replication within even the larger, basin-wide areas. Instead, replication was introduced at the level of each station, and therefore the statistical populations in question coincide roughly with the immediate area of each sampling station. This proved to be particularly fortuitous since subsequent data has indicated that substantial biological changes can and do occur in the offshore at spatial scales considerably smaller than whole lake basins.

1.2 Objectives of study

The primary goal of this study was to determine if GLNPO's DQO is being met with the current level of sampling effort in the benthic program. Specifically, the goals of this study were several fold:

- To assess whether inter-site variability at offshore stations precludes replication on a basin-wide basis;
- To determine the minimum detectable differences under the current sampling regime;
- To determine the sample sizes required to meet the DQO;
- To assess the current DQO in relation to the magnitude of variability seen in benthos data.

In addition, different possible interpretations of the GLNPO DQO, and its general suitability for biological data will be discussed.

2 GLNPO's Data Quality Objectives

2.1 DQO for GLNPO water quality survey

In order to assess lake health using data generated from the benthic program, or from any aspect of the water quality survey, sufficient data quality must be obtained to permit detection of 'significant' changes in these variables. For the purposes of the water quality surveys, GLNPO has defined a significant change as a 20% difference between current and historical measurements, made for a particular variable in a particular lake during a particular season. Data quality should be sufficient for there to be an 80% chance of detecting such a change at the 90% confidence level.

2.2 Application of DQO to benthic data

As currently formulated, GLNPO's DQO might not be strictly applicable to the benthic program since, as noted above, data generated from nearshore stations were only intended to be representative of local conditions, while even offshore stations might capture too much site-specific variation to be usefully pooled to represent broader, basin- or lake-wide areas. In this case, changes might be more profitably assessed on a station by station basis, rather than on a whole-lake, or even basin-wide, basis.

Additionally, application of the DQO requires an explicit statement of how the benthos data are to be used to assess differences between current and historical measurements. As detailed in the next section, 'historical measurements' can be interpreted in a number of different ways. While perhaps these differences might seem to be largely semantic, the ramifications of the different interpretations can dramatically change the fundamental questions being asked by the monitoring program, and the statistical techniques used to answer those questions. By referring to difference between current and 'historical' measurements, rather than, say, year to year differences, the current DQO implies that historical measurements refer to the pooling of all past data. In this case, changes in a given variable in the past would contribute to variability in the historical measurements, and hinder the detection of further changes without continual increases in sample size. If changes are not unidirectional (e.g., are cyclical), then pooling historical data could completely preclude detection of further changes. Two other possible interpretations of the DQO include the ability to detect differences between any two (or more) years, and the ability to detect directional changes (trends). The statistical implications of each of these interpretations is discussed in the next section.

2.3 Choice of response variables

While a precise definition of historical measurements is necessary before the data can be assessed for its ability to satisfy the stated DQO, so is a precise definition of the response variable. Unlike most chemical data, biological data, including that generated from the benthos program, is multivariate, and therefore offers a number of potential response variables. For instance, changes in the densities of individual species can be assessed, potentially limited to either specific indicator species or to dominant species. In this case, determination of the ability of the data to satisfy the DQO would require multiple analyses, one for each species of interest. The total density of the benthic community could be also be used, or the total densities of individuals within more broadly defined taxonomic categories (e.g., oligochaetes, chironomids, etc.). Alternatively, community-level metrics, such as diversity or species richness, could be used. Finally, specific indicators of benthic community structure, such as the Milbrink (Milbrink, 1983), Goodnight and Whitley (Goodnight and Whitley, 1960), or Brinkhurst (Brinkhurst, 1967) oligochaete indices, could be employed. Presumably the variables to be subjected to the DQO criterion should be ones which are conceptually tractable, and for which changes would have some understood ecological meaning.

3 Possible Interpretations of GLNPO's DQO

Determination of the adequacy of the current benthos sampling regime in satisfying GLNPO's DQO involves an assessment of the precision in estimating the mean (μ) of the statistical population in question. The precision of a given statistical estimate is affected by the natural variation, or variance (σ^2), of the population under study, and the sampling variability. Since natural variation is not under investigator control, increases in precision can only be effected through decreases in sampling variability, which is mainly accomplished by increasing sample size (n), although theoretically improvements in sampling methodology can also result in reduced sampling variability. If a desired precision in estimating a parametric mean is known in advance, as is the case with a DQO, and the desired probability of attaining that precision for a given α is also specified, then the number of samples needed to achieve the desired precision can be calculated according to the statistical test to be used.

GLNPO's DQO addresses the detection of a change in a population mean, relative to an historical value. The precision with which a change can be detected will be a function not only of the number of observations in the current sample, but also of the number of observations making up the historical sample. The adequacy of a given sample size to detect such a change will therefore depend upon the constitution of the historical sample, and this in turn depends upon the exact definition of the 'historical sample', and on the statistical test which is to be used to assess differences between it and the current sample. A number of possible interpretations of the DQO exist, and three of the most likely are outlined below, along with the statistical considerations involved in their assessment.

3.1 Detection of differences between current year and all previous years combined

In the simplest scenario, the current sample is compared to all past samples pooled. This corresponds to the most literal interpretation of GLNPO's DQO, and it assumes that all past samples estimate the same population mean, i.e., that there have been no changes in the variable of interest prior to the current year. In this case, the appropriate formula for determining sample size is:

$$n = \frac{2s_p^2\left(t^2_{\alpha(2),2(n-1)}\right)^2 F_{\beta(1),(n-1,)\nu}}{d^2}$$

where: s_p^2 = sample estimate of pooled population variance; and

d = the minimum detectable difference specified by the DQO.

The assumption that there have been no changes in past years would require statistical testing to assess, and therefore a *de facto* testing for interannual differences between all years would be required under this interpretation of the DQO, whether or not that testing were, strictly speaking, being used to assess differences of interest (i.e., differences between the current year and all previous years combined). Furthermore, the assumption that all historical data are statistically the same contains an element of self-contradiction, since by definition each current year's data becomes historical data with the collection of the next year's data.

3.2 Detection of changes between any year

A second, and more likely, interpretation of the DQO is that the detection of *any* interannual differences that have occurred over the course of the monitoring program is of interest. In this case, each year's data would constitute a different 'treatment', and a one-way ANOVA, or equivalent, would be used to assess differences between 'treatments' (i.e., years). The sample size necessary to detect a specified difference (δ) in mean values can be calculated using the following equation:

$$\phi = \sqrt{\frac{n\delta^2}{2ks^2}}$$

where: k = number of treatments (in this case years)
n = number of observations in each treatment
s^2 = within groups variance
δ = the minimum detectable differ ence, specified by the DQO

ϕ is a parameter related to the power of the performed test. This equation is solved iteratively, inputting successive estimates of n until a ϕ corresponding to the desired 1-β is achieved.

Alternatively, the following equation can be used to test for the minimum detectable difference with a specified n:

$$\delta = \sqrt{\frac{2ks^2\phi^2}{n}}$$

Under this scenario, the target minimum detectable difference of 20% would apply to changes seen between any two years, even non-consecutive years. Of course with an analysis of variance procedure, exactly *which* treatments (in this case years) are different from which others are not specified; what can be concluded from the analysis is only whether or not *any* of the treatments are different from any other treatments.

3.3 Detection of trends over time

A third possible interpretation of the DQO would be to enable detection of a trend in the variable of interest, in other words, to test for a change over time. This can be done in one of two ways: through the use of a regression analysis, with the variable of interest regressed against time (i.e., year), or by using a correlation analysis, again with the variable of interest correlated with time. In the first case, the *magnitude* of change over time can be assessed, specifically by examination of the regression coefficient, *b*. Simple linear regression, however, has at least two drawbacks in the context of the current application. First, it assumes a cause and effect relationship between the independent and dependent variables. Strictly speaking, it is unlikely that time in and of itself would be the causative factor in increases or decreases in benthos densities. Second, and more importantly, it assumes that the dependent variable (e.g., benthos density at a site) has a *constant*, *linear* relationship to the independent variable (time). If the detection of *any* change in a variable over time is of interest, and not just a strictly linear one, then linear regression analysis is an inappropriate tool.

Alternatively, correlation analysis can be used to assess whether changes in the magnitude of one variable are associated with changes in the magnitude of a second variable. In correlation analysis, no cause and effect relationship is assumed. However, in simple linear correlation analysis, a linear relationship between the two variables of interest is assumed. This restriction can be circumvented by use of a rank-based correlation procedure, such as the Spearman rank correlation procedure. With this procedure, data are converted to ranks prior to conducting the analysis, so that only trends in the two variables are assessed, and not quantitative changes of one in relation to the other variable. So in this case, the *magnitude* of the change is not assessed, rather the *strength of association* between the two variables is. For this reason, the GLNPO DQO as currently formulated cannot be applied to an analysis of this nature, since the DQO addresses a specific magnitude of change (i.e., a 20% change over some historical value).

For the purposes of this report, it was assumed that the GLNPO DQO requires the ability to detect a 20% change in a variable between any two (or more) years for which data are available. The approach adopted here, therefore, follows that outlined above in section 3.2.

4 Overview of Current Benthos Methods

In 1999, the most current year for which full benthos data is available, a total of 53 stations were visited during the summer survey (Figure 1). Between seven (Lake Erie) and fourteen (Lake Michigan) stations were visited in each lake. At each station, three replicate samples were collected, using a Ponar grab sampler. Samples were sieved through a 500-μm sieve in the field, preserved, and transported to the laboratory. In the laboratory, all animals were removed from any remaining sediment under a dissecting microscope and sorted by major taxonomic group. All oligochaetes and chironomids were mounted and identified under a compound microscope; all other invertebrates were identified and counted under a dissecting microscope. Counts were converted into areal units (#/m^2) by multiplying by 19.12, a factor which takes into account the area sampled by the Ponar.

5 Approach

In all cases, data from 1999 were used, since this is the most recent year for which complete benthos data are available. For the purposes of this study, offshore stations are defined as those with a depth > 64 m. This is a slightly narrower definition than the one originally employed in the design of the benthic survey, and was adopted largely because of the high percentage of stations with depths within a few meters of 50 m, the original demarcation between near- and offshore stations. The variables considered for offshore stations in this study were areal densities (#/m^2) of the total benthos community, the total oligochaete community (excluding fragments, but including immatures), and the two common profundal organisms *Diporeia* and *Stylodrilus*. Because of greater species richness, additional variables were examined at nearshore stations. These included total areal densities of the Naididae, the Tubificidae, the Chironomidae, and the Sphaeriidae. Sites were excluded from analysis where numbers were extremely small, or where organisms were completely absent from one or more replicate.

Since the original rationale of the sampling design was that offshore stations within a lake would be estimating the same statistical population, an initial analysis was undertaken to assess within lake differences in each of the above four variables. These potential differences were assessed using a one way analysis of variance (with stations as a random factor). Where assumptions of homoscedasticity or normality were not met, a Kruskal-Wallis one way analysis of variance on ranks was performed.

Minimum detectable differences were determined for all variables tested at all sites using the equation in section 3.2. Data were not tested for homoscedasticity or normality prior to computation of minimum differences, nor were any transformations used. It is recognized that in some cases these assumptions were probably not met. However, given the number of analyses, it was not deemed feasible to assess the need for transformations in each case, particularly given the robustness of a one way ANOVA, the statistical test upon which most calculations were based. Minimum detectable differences were determined

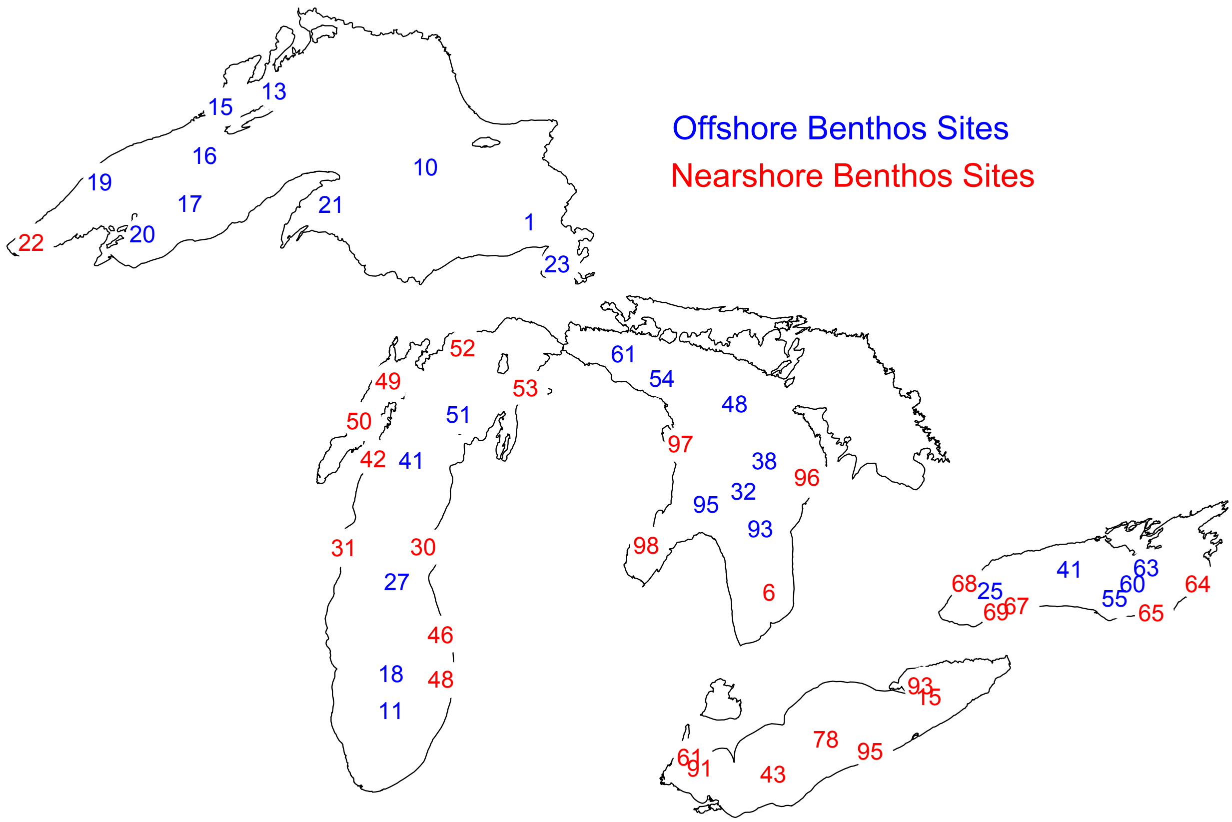

Fig. 1. Locations of GLNPO's water quality survey (WQS) benthos sampling stations for summer, 1999 survey. Sites considered offshore are indicated in blue; sites considered nearshore are indicated in red. .

for n (i.e., number of replicates taken) = 3, 4 and 5, and for k (i.e., number of years compared) = 2 through 9. In all cases, β was taken to equal 1-0.80 and α was taken to equal 0.05. The most recent GLNPO DQO assumes an α = 0.10; however, tabled values for ϕ given α = 0.10 and varying degrees of freedom could not be found, so the more commonly used α of 0.05 was used instead. Therefore, estimates of minimum detectable differences will be somewhat more conservative than are required by the current DQO. General trends, however, should not be affected by the difference in α.

The adequacy of the current sampling regime in satisfying the GLNPO DQO was examined assuming that any between-year differences are of interest. Natural variation at each site was estimated using sample variance calculated from 1999 data. The number of replicates required to satisfy the stated GLNPO DQO was calculated for each variable according to the equation in section **3.2**, using the mean value of the variable in 1999 as the value in relation to which a 20% change should be detectable, and assuming a comparison of 5 years, α = 0.10 and β = 1-0.80.

6 Results

6.1 Within Lake Variation in Offshore Sites

ANOVA analyses were conducted on offshore sites within each lake to determine if

Table 1. One way analysis of variance results examining between site differences in Lakes Huron, Michigan, Ontario and Superior for offshore stations. For all tests, $\alpha = 0.05$. Where assumptions of normality or homoscedasticity were not met, the Kruskall-Wallis one way analysis of variance on ranks was used.

LAKE	Σ Benthos		Σ Oligochaetes		*Diporeia*		*Stylodrilus*	
	F	P	F	P	F	P	F	P
HURON	5.08	0.01	4.35	0.01	5.32	0.01	17.6*	0.01
MICHIGAN	94.6	<0.001	24.7	<0.001	55.2	<0.001	14.2	<0.001
ONTARIO	10.6*	NS	13.5*	0.02	38.1	<0.001	15*	0.01
SUPEROR	28.7	<0.001	26*	0.002	26.5*	0.002	23.3*	0.01

* Kruskal-Wallis One Way Analysis of Variance on Ranks used: statistic = H.

these sites should serve as replicates within each lake. Variables tested were densities of total benthos, total oligochaetes, *Diporeia* and *Stylodrilus*. For all variables tested, with the sole exception of total benthos densities in Lake Ontario, significant differences were found between sites at $\alpha = 0.05$ (Table 1). Therefore all sites were considered separately in subsequent analyses.

6.2 Comparison of variability between offshore and nearshore sites

To determine if within site variability was related to depth, i.e., if nearshore sites were more or less variable than offshore sites, the standard deviation of total benthos estimates for individual sites was plotted against depth for all sites. The more extreme standard deviations were associated with shallower sites, although not all shallow sites exhibited a large standard deviation (Figure 2). Since the GLNPO DQO is formulated in terms of percent, rather than absolute, change, though, high standard deviations would not necessarily result in large required sample sizes if those standard deviations were associated with high values of the variable of interest. Benthos densities do in fact tend to be higher at shallower sites, and this might counteract the higher variances seen at shallow sites.

The coefficient of variation:

$$CV = \frac{SD}{\bar{X}}$$

provides a measure of the variance relative to the mean, and when this was plotted against depth, no relationship was seen (Figure 2). This indicates that the higher standard deviation of some shallow sites was due to higher total numbers.

When the natural logarithm of the sample size required to satisfy the DQO for the estimation of total benthos at each site was plotted against site depth, an apparent positive relationship was seen (Figure 3). This relationship was just statistically significant at $\alpha = 0.10$, as determined by least squares regression ($F = 2.77$, $P = 0.10$). A more significant relationship was found between the sample size required to estimate total numbers of oligochaetes and depth ($F = 5.59$, $P = 0.02$). In both cases depth explained a relatively low percentage of the variance in required sample size (total benthos: $R^2 = 0.03$; total oligochaetes: $R^2 = 0.09$). This suggests that, on the whole, deeper sites tend to require a greater degree of replication to enable detection of a 20% change for these variables than

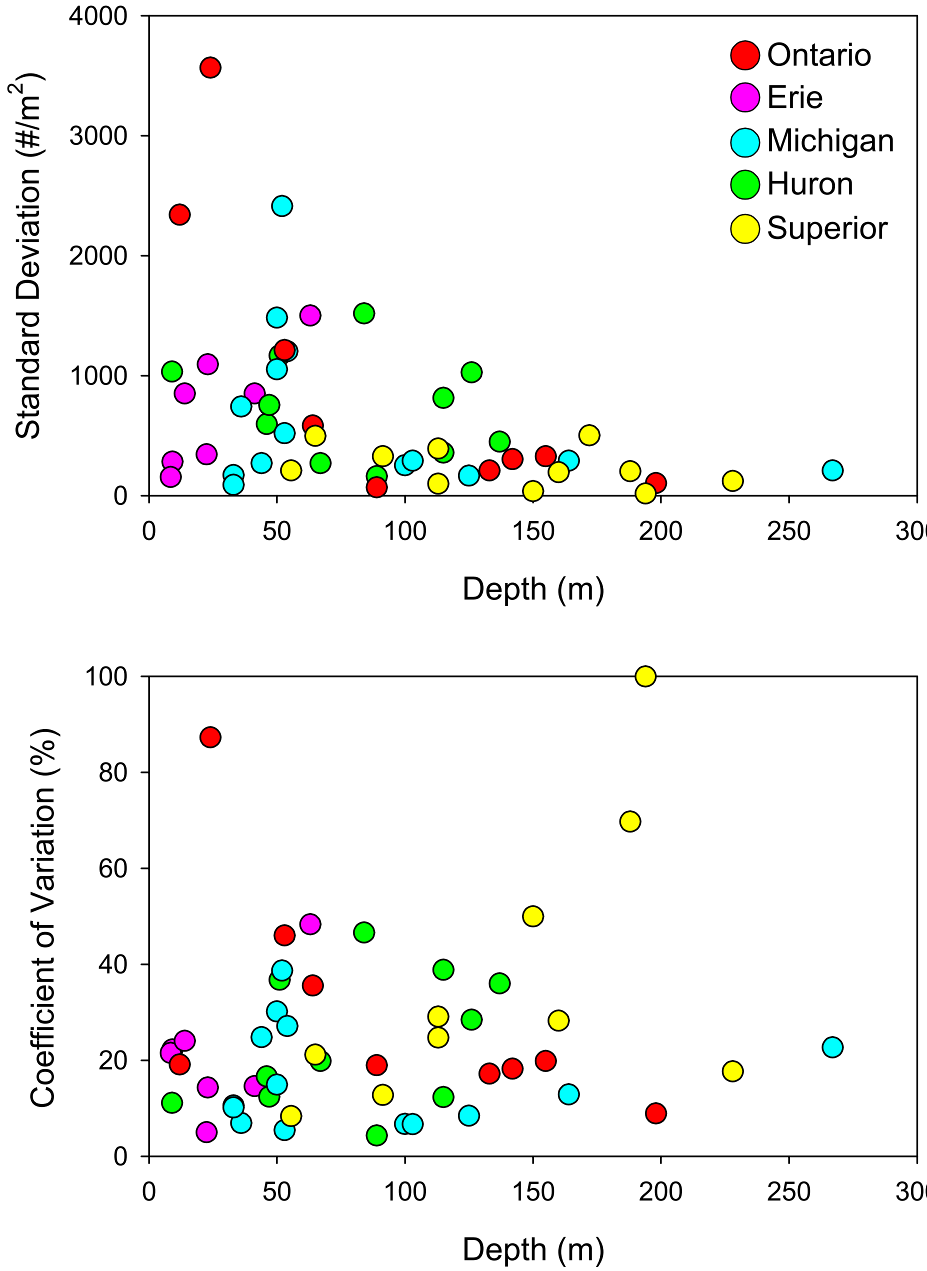

Figure 2. Standard deviation (top panel) and coefficient of variation (bottom panel) of estimates of total benthos densities at each site, graphed against site depth. Variance estimates based on 1999 values.

do shallower sites, although this tendency is not marked. As noted, this is due in large part to the smaller densities of organisms seen at deeper sites. Such a relationship was not found, however, between sample sizes required to satisfy the DQO for densities of *Diporeia* and station depth (Figure 3). The probable reason for this is that, unlike total benthos and total oligochaete densities, *Diporeia* numbers tend to increase with depth.

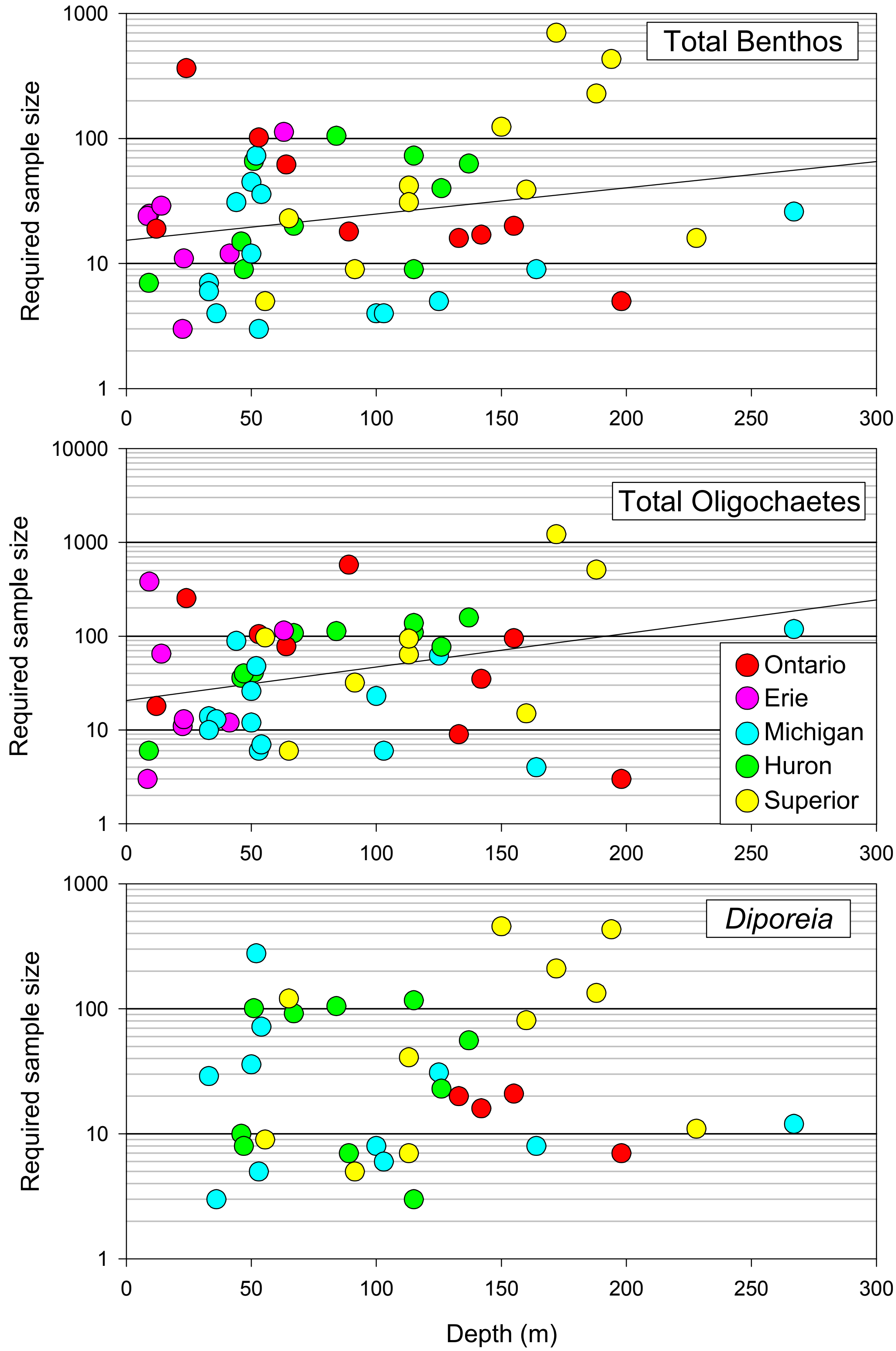

Figure 3. Relationship between sample size required to satisfy DQO and site depth for estimation of total benthos, total oligochaete and *Diporeia* areal densities. In all cases, $k = 5$, $\alpha = 0.10$ and $\beta = 1 - 0.80$.

6.3 Minimum detectable differences for interannual comparisons at offshore sites

Average lake-wide minimum detectable differences for the current level of replication, assuming a comparison of 5 years of data and $\alpha = 0.05$, were less than 1,000 individuals/m² in all cases except in Lake Huron, where the average minimum detectable difference for *Dioreia* was 1,692/m² and for the total benthos was 2,343/m² (Table 2). This is reflective of the generally higher variances, and hence higher minimum detectable differences, seen at the offshore sites in that lake compared to the other lakes. Average minimum detectable differences for the four variables were broadly similar in Lakes Ontario and Michigan, and somewhat lower in Lake Superior, due to the lower densities of organisms in that lake.

Since the GLNPO DQO is stated in terms of a percent, rather than an absolute, change, a more relevant measure of the adequacy of the current sampling program in fulfilling the requirements of the DQO is minimum percent detectable difference. In most cases, the

Table 2. Maximum, minimum and average minimum detectable differences (#/m²) and percent detectable differences for offshore ($z > 64$ m) stations in Lakes Michigan, Huron, Ontario and Superior, under the current sampling regime. In all cases, $n = 3$, $k = 5$, $\alpha = 0.05$ and $\beta = 0.80$.

LAKE	Minimum Detectable Difference			Percent Detectable Difference		
Variable	Max	Min	Avg	Max	Min	Avg
MICHIGAN						
Diporeia	1,144	314	794	89%	34%	52%
Stylodrilus	1,109	157	580	617%	46%	188%
Σ Oligochaetes	1,331	157	589	178%	27%	88%
Σ Benthos	1,040	595	864	81%	24%	41%
HURON						
Diporeia	3,055	491	1,692	175%	22%	108%
Stylodrilus	2,289	104	651	310%	21%	165%
Σ Oligochaetes	2,093	68	965	206%	14%	152%
Σ Benthos	5,407	579	2,343	166%	15%	95%
ONTARIO						
Diporeia	1,020	343	741	72%	40%	61%
Stylodrilus	1,570	39	443	168%	62%	123%
Σ Oligochaetes	1,954	39	443	392%	21%	143%
Σ Benthos	2,084	245	951	127%	32%	71%
SUPERIOR						
Diporeia	1,513	68	437	356%	30%	168%
Stylodrilus	1,396	104	539	592%	44%	236%
Σ Oligochaetes	1,454	39	531	617%	35%	253%
Σ Benthos	1,791	68	855	432%	45%	169%

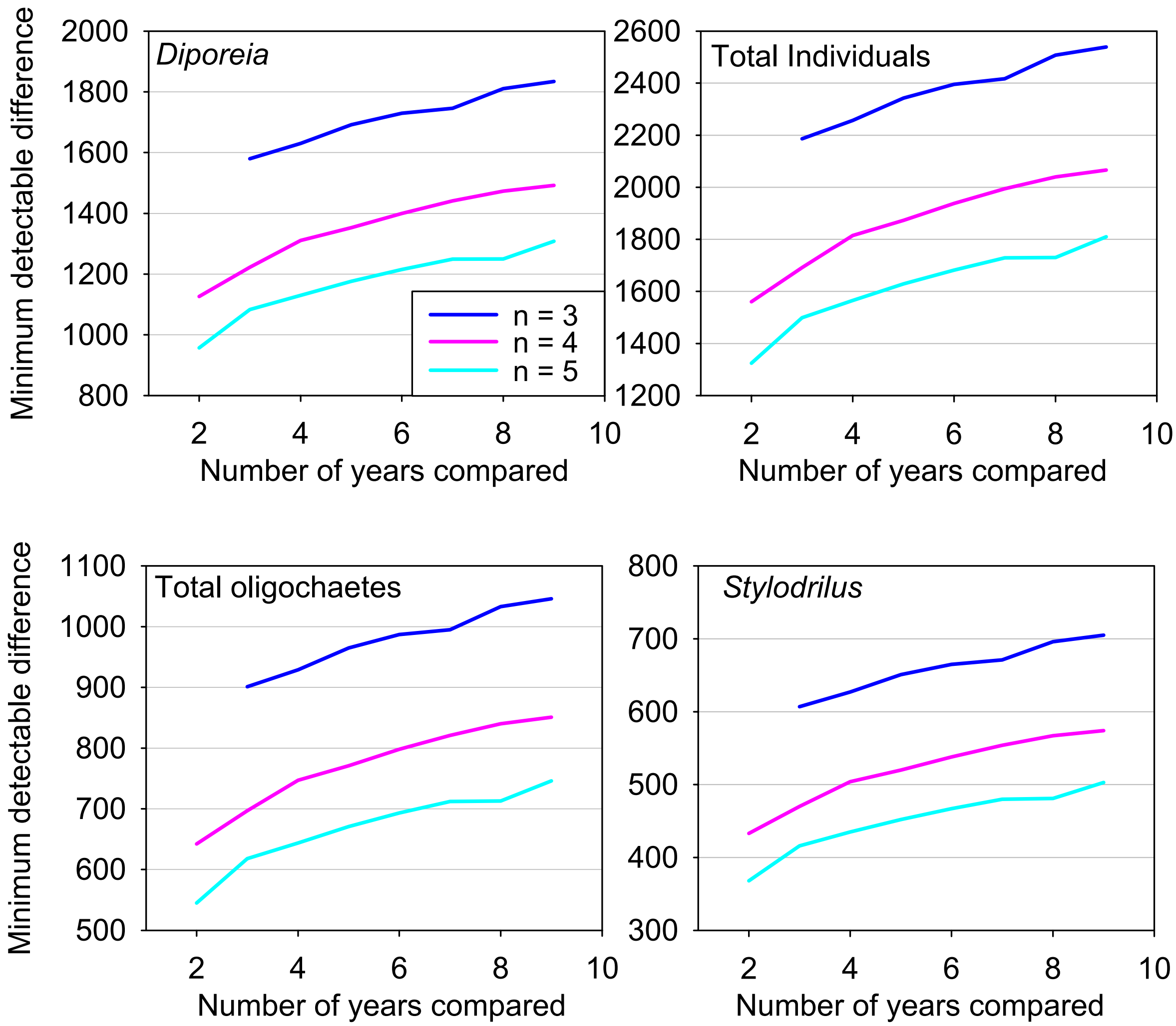

Figure 4. Average minimum detectable differences in areal densities of *Diporeia*, total benthos, total oligochaete and *Stylodrilus* at offshore sites in Lake Huron for different numbers of replicates (*n*) when comparing 2 - 9 years of data. Variance estimates are based on 1999 data.

smallest percent difference that the current sampling regime permits detection of was substantially larger than 20%, reaching several hundred percent in a large number of cases (Table 2). In only two cases was the current level of replication adequate to permit detection of a 20% change: given estimates of variance from 1999, current levels of replication permit detection of 14% and 15% changes in total oligochaete and total benthos densities, respectively, at station HU 93 in Lake Huron. Average minimum percent detectable differences were highest in Lake Superior; this was due to not to higher variance, but rather to lower overall densities of organisms in the lake. Thus, while in general smaller differences are able to be detected in Lake Superior, the low densities seen in the lake counteract this to make minimum percent detectable differences larger than in the other lakes.

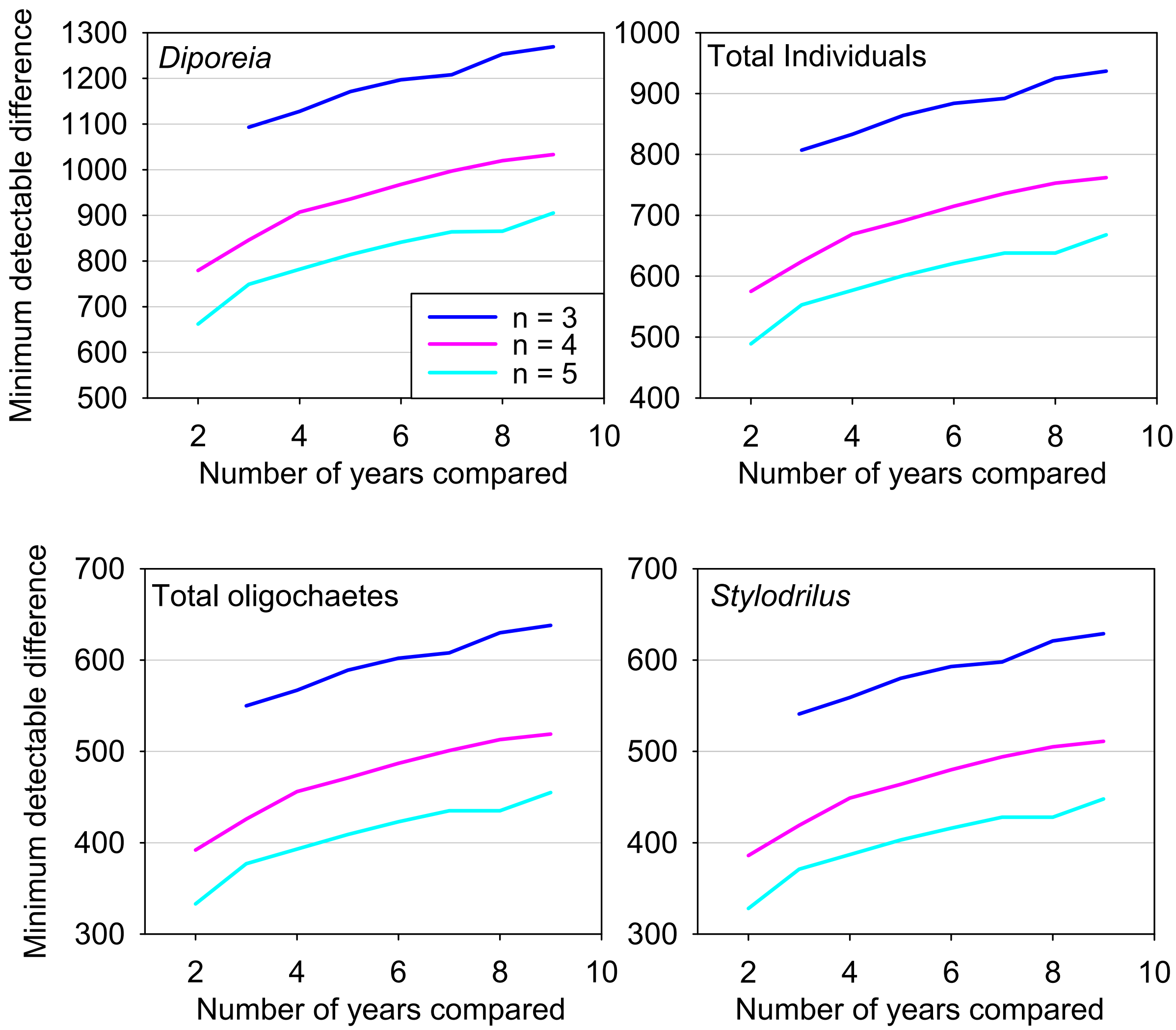

Figure 5. Average minimum detectable differences in areal densities of *Diporeia*, total benthos, total oligochaete and *Stylodrilus* at offshore sites in Lake Michigan for different numbers of replicates (*n*) when comparing 2 - 9 years of data. Variance estimates are based on 1999 data.

Minimum detectable differences were calculated for interannual comparisons in which between 2 to 9 years of data were compared, and with 3, 4 and 5 replicates per site, for α = 0.05. For any given level of replication, minimum detectable differences increase as the number of years in the comparison increases (Figures 4 - 7). In general, increasing the number of replicates from 3 to 4 resulted in a decrease in minimum detectable differences of approximately 17-23%, while increasing replication to 5 further reduced minimum detectable differences by approximately 10-15%. In absolute terms, taking 4 replicate samples at each site reduced the minimum detectable difference by an average of 170 individuals/m^2 for the four variables considered (range: 88 - 470/m^2), while 5 replicate samples per site reduced the minimum detectable difference by an average of 259/m^2 (range: 133 - 714/m^2), compared to 3 replicates (Table 3).

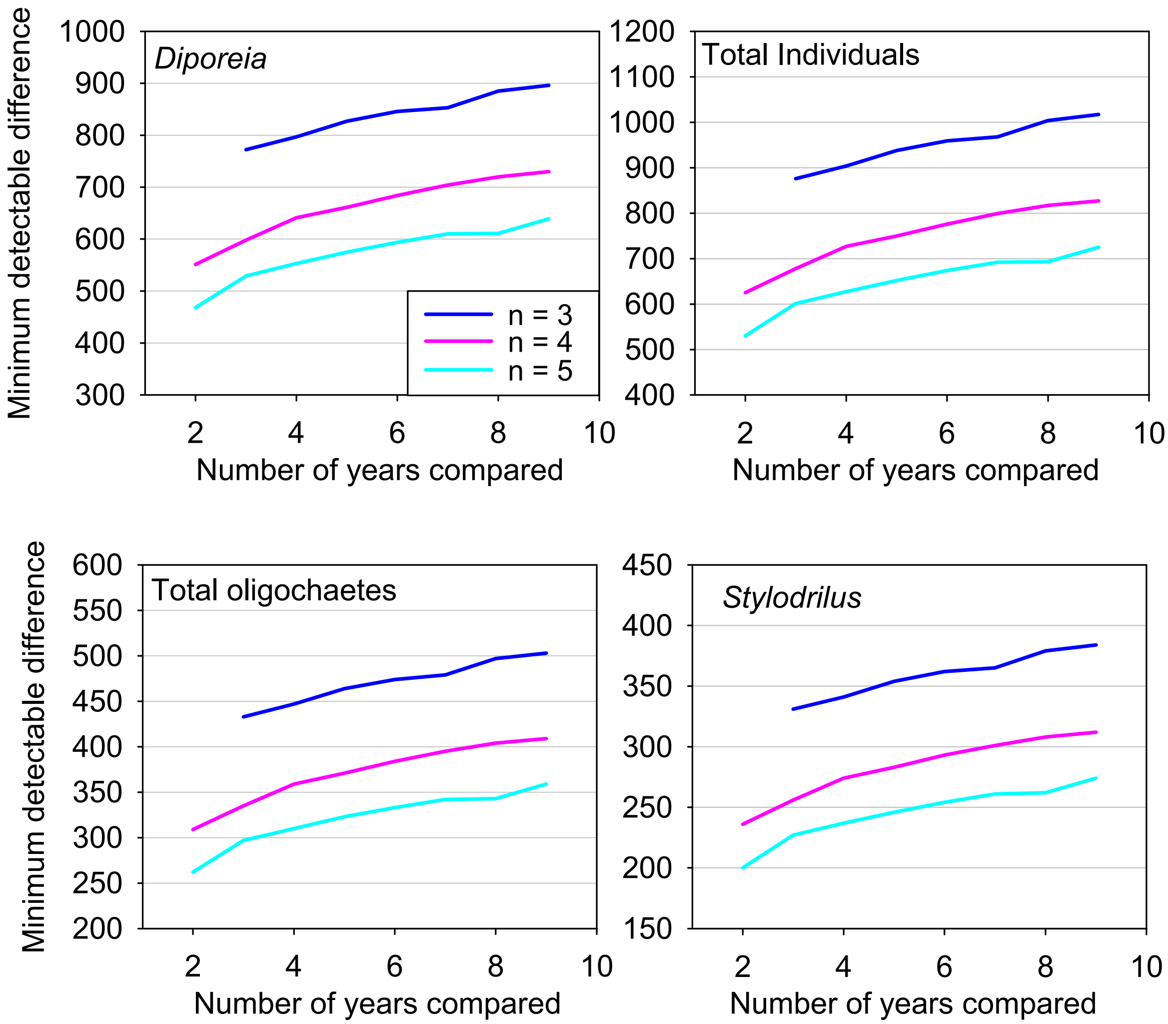

Figure 6. Average minimum detectable differences in areal densities of *Diporeia*, total benthos, total oligochaete and *Stylodrilus* at offshore sites in Lake Ontario for different numbers of replicates (*n*) when comparing 2 - 9 years of data. Variance estimates are based on 1999 data.

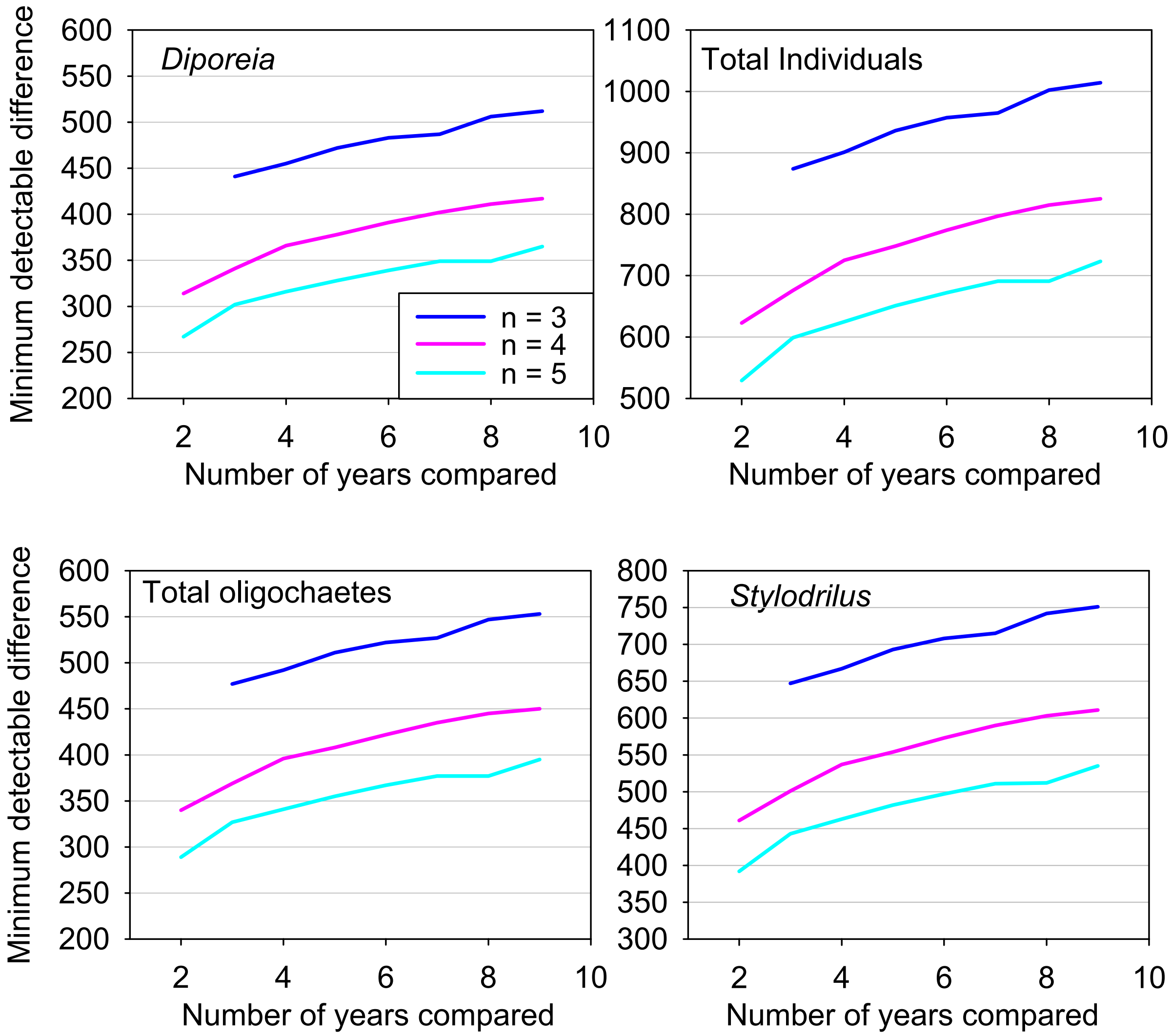

Figure 7. Average minimum detectable differences in areal densities of *Diporeia*, total benthos, total oligochaete and *Stylodrilus* at offshore sites in Lake Superior for different numbers of replicates (*n*) when comparing 2 - 9 years of data. Variance estimates are based on 1999 data.

Table 3. Estimated average minimum detectable differences (#/m²) for off-shore (z > 64 m) stations in Lakes Michigan, Huron, Ontario and Superior, for samples sizes of 3, 4 and 5. In all cases, $k = 5$, $\alpha = 0.05$ and $\beta = 1-0.80$.

LAKE	Sample Size		
Variable	3	4	5
MICHIGAN			
Diporeia	794	635	552
Stylodrilus	580	464	403
Σ Oligochaetes	589	471	409
Σ Benthos	864	691	601
HURON			
Diporeia	1,692	1,353	1,176
Stylodrilus	651	520	452
Σ Oligochaetes	965	771	671
Σ Benthos	2,343	1,873	1,629
ONTARIO			
Diporeia	741	593	515
Stylodrilus	594	474	413
Σ Oligochaetes	443	354	308
Σ Benthos	951	760	661
SUPERIOR			
Diporeia	437	349	304
Stylodrilus	539	431	375
Σ Oligochaetes	531	424	369
Σ Benthos	855	684	595

6.4 Required sample sizes for offshore sites

The number of replicates needed to detect a 20% change in the four variables tested, assuming a five year comparison, ranged from 3 for the site in Lake Huron noted above (HU 93), to 702 to detect a 20% change in total benthos densities at site SU 16 in Lake Superior (Figure 8). The median number of replicates needed to meet the DQO for all variables was 39, while fully 80% of the variable/ site combinations tested would require at least 12 replicates to satisfy current the DQO. However, since each variable is not sampled independently at each site, the number of replicate samples required on a site by site basis would be determined by the variable requiring the greatest number of replicates at that site to satisfy the DQO. Examined on this basis, the median number of replicates at offshore sites needed to meet the DQO was estimated to be 97, while 80% of sites would require 19 replicates or more. These estimates obviously represent an impractical degree of replication, both from the standpoint of ship time required to amass such a number of samples, and from the standpoint of time required for sample analysis.

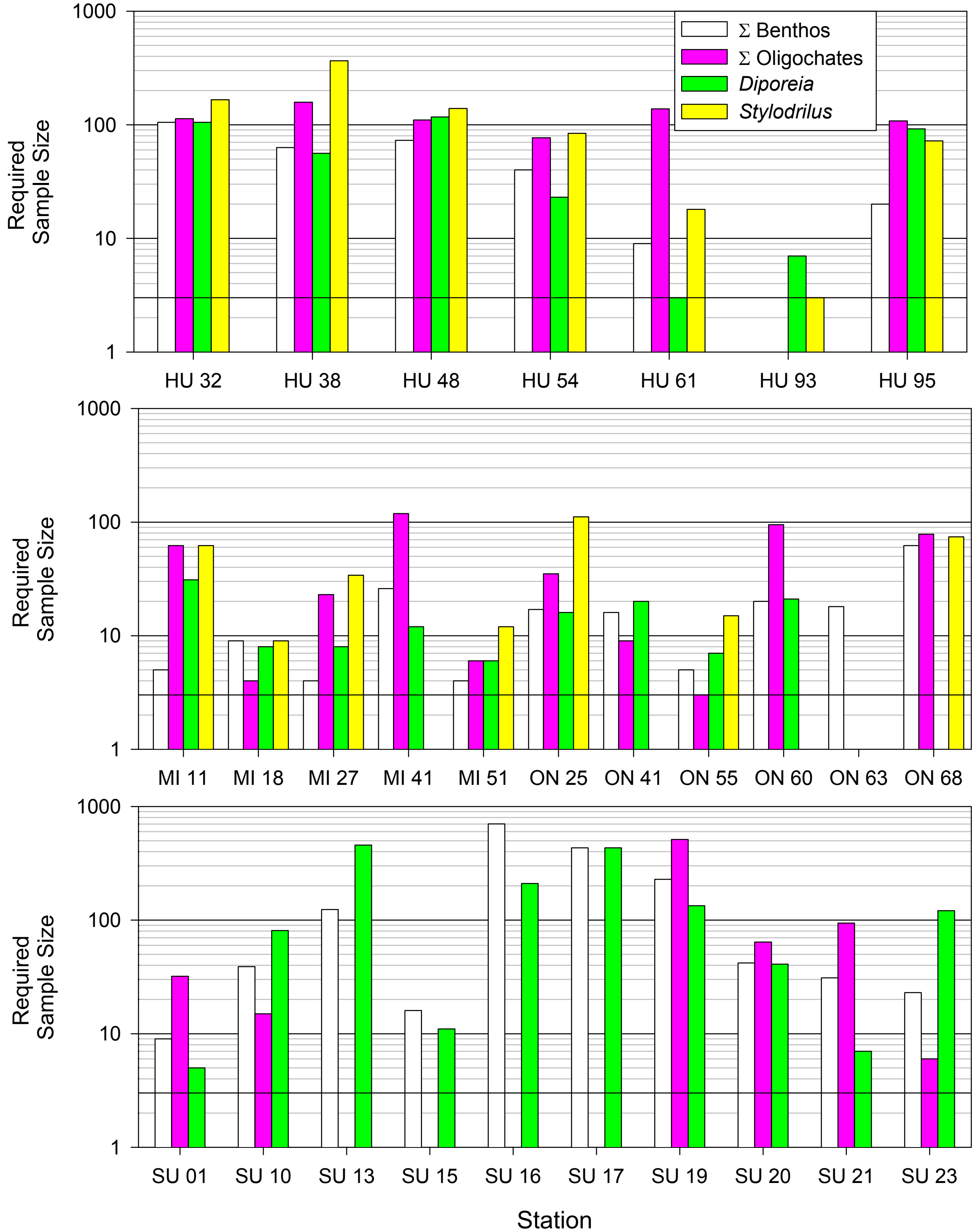

Figure 8. Sample sizes required at offshore sites to detect a 20% change in total benthos, total oligochaetes, *Diporeia* and *Stylodrilus* areal densities when comparing 5 years, assuming α = 0.10 and β = 1 - 0.8. Variance estimates and initial densities are based on 1999 values. Reference line represents n = 3.

6.5 Minimum detectable differences for interannual comparisons at nearshore sites

Minimum detectable differences were calculated for each nearshore site, assuming a comparison of five years and $\alpha = 0.05$, for the following six variables: areal densities of total benthos, total oligochaetes, total Tubificidae, total Chironomidae, *Diporeia* and Sphaeridae. In only five cases was the current level of replication sufficient to detect a 20% difference when comparing five years of data (Figure 9). Most values for Lakes Erie, Huron and Michigan were between 50 and 100% (medians = 65%, 59% and 79%, respectively), while minimum detectable differences were somewhat higher in Lake Ontario (median = 166%). Of the three nearshore stations in the latter lake, two exhibited substantial variability in all variables examined. No substantial differences in minimum detectable differences were apparent between the different variables.

6.6 Required sample sizes for nearshore sites

The number of replicates needed to detect a 20% change in the six variables tested, assuming a five year comparison and $\alpha = 0.10$, ranged from three for several variables at three sites in Lakes Erie and Michigan, to 837 replicates to estimate total chironomids at a station in Lake Ontario (Figure 10). Overall, a median sample size of 25 was required when all variables at all sites were considered, with 80% of site/variable combinations requiring nine replicates or more. When just the maximum required sample size at each site for all six variables was considered, which is a better indication of the required site-wise level of replication, the median number of replicates needed to satisfy the DQO at all sites was 77, with 80% of the sites requiring 51 replicates or more for the DQO to be met for all variables tested. Given the greater number of variables under consideration here, this is a more rigorous test of the sampling program than was used for offshore stations. Considering just those variables that were also assessed in the offshore sites, namely total benthos, total oligochaetes and *Diporeia* (*Stylodrilus* does not appear in substantial numbers at shallower sites), the median number of replicates needed to satisfy the DQO at each site was 38, with 80% of the sites requiring at least 13 replicates. These numbers are lower than those seen in the offshore sites. On average, then, nearshore sites would require less replication to satisfy GLNPO's DQO, compared to offshore sites. Even so, the level of replication required at nearshore sites for acceptable estimates of total benthos, total oligochaetes and *Diporeia* densities, given the current DQO, are in the vast majority of cases infeasible from a practical standpoint.

7 Relation of DQO to 1997-2001 *Diporeia* data

To determine the relation of the GLNPO DQO to a benthic variable of ecological interest, and for which data from 1997-2001 are available, changes in the density of *Diporeia* were examined. The amphipod *Diporeia* has historically been one of the most abundant and widespread benthic organisms in the Great Lakes (Dermott and Corning, 1988) and is an important link in the Great Lakes food chain, feeding on pelagic-derived detritus (Gardner *et al.*, 1990) and in turn providing an important food source for many fish species (Scott and Crossman, 1973). Recently its numbers have been declining in significant portions of its range in the Great Lakes, with potentially huge consequences for community structure and energy flow through the Great Lakes food web.

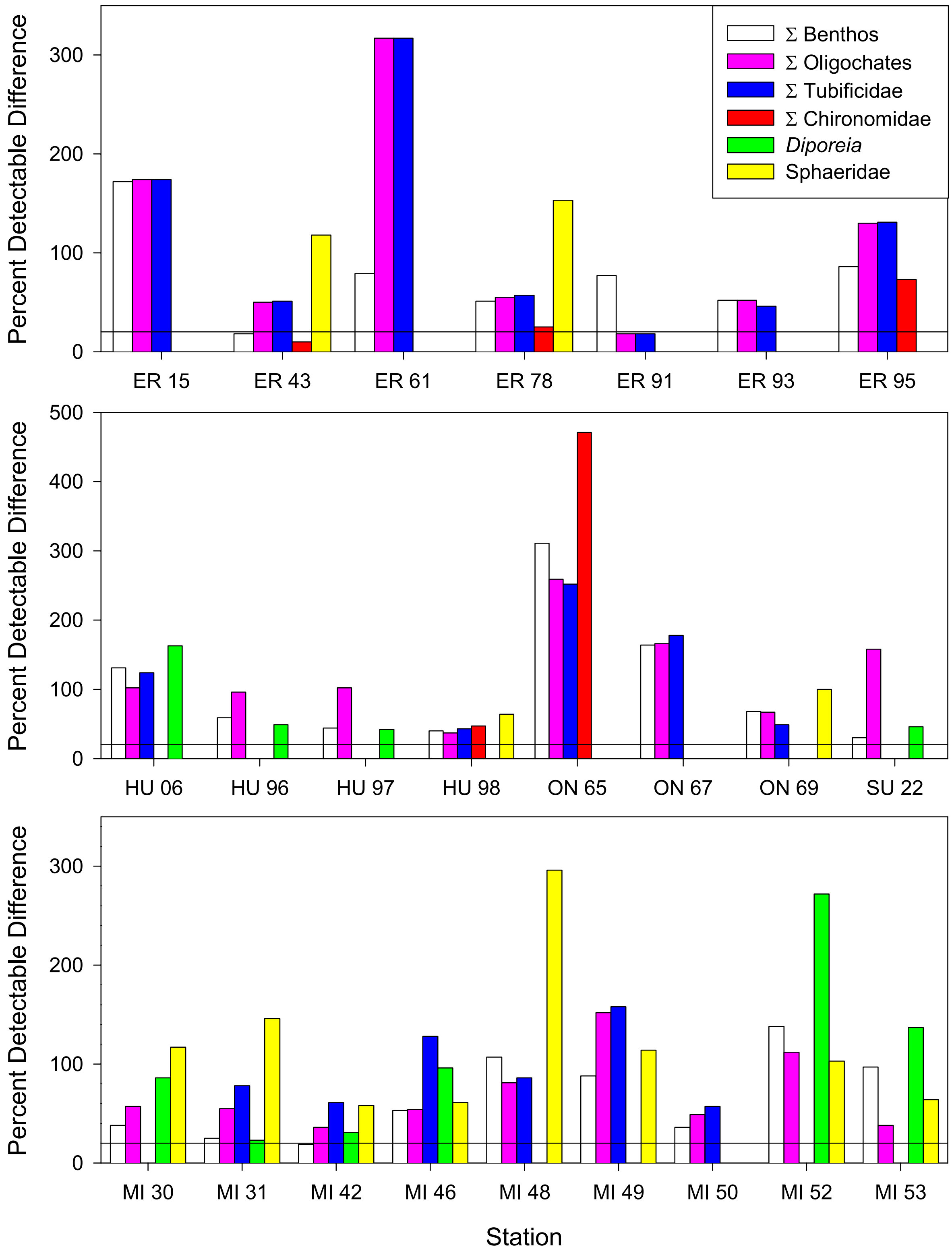

Figure 9. Minimum detectable differences, as a percent, for estimates of total benthos, total oligochaete and total Tubificidea, total Chironomidae *Diporeia* and Sphaeridae areal densities. In all cases, $k = 5$, $\alpha = 0.05$ and $\beta = 1 - 0.80$. Reference line represents current data quality objective of 20%

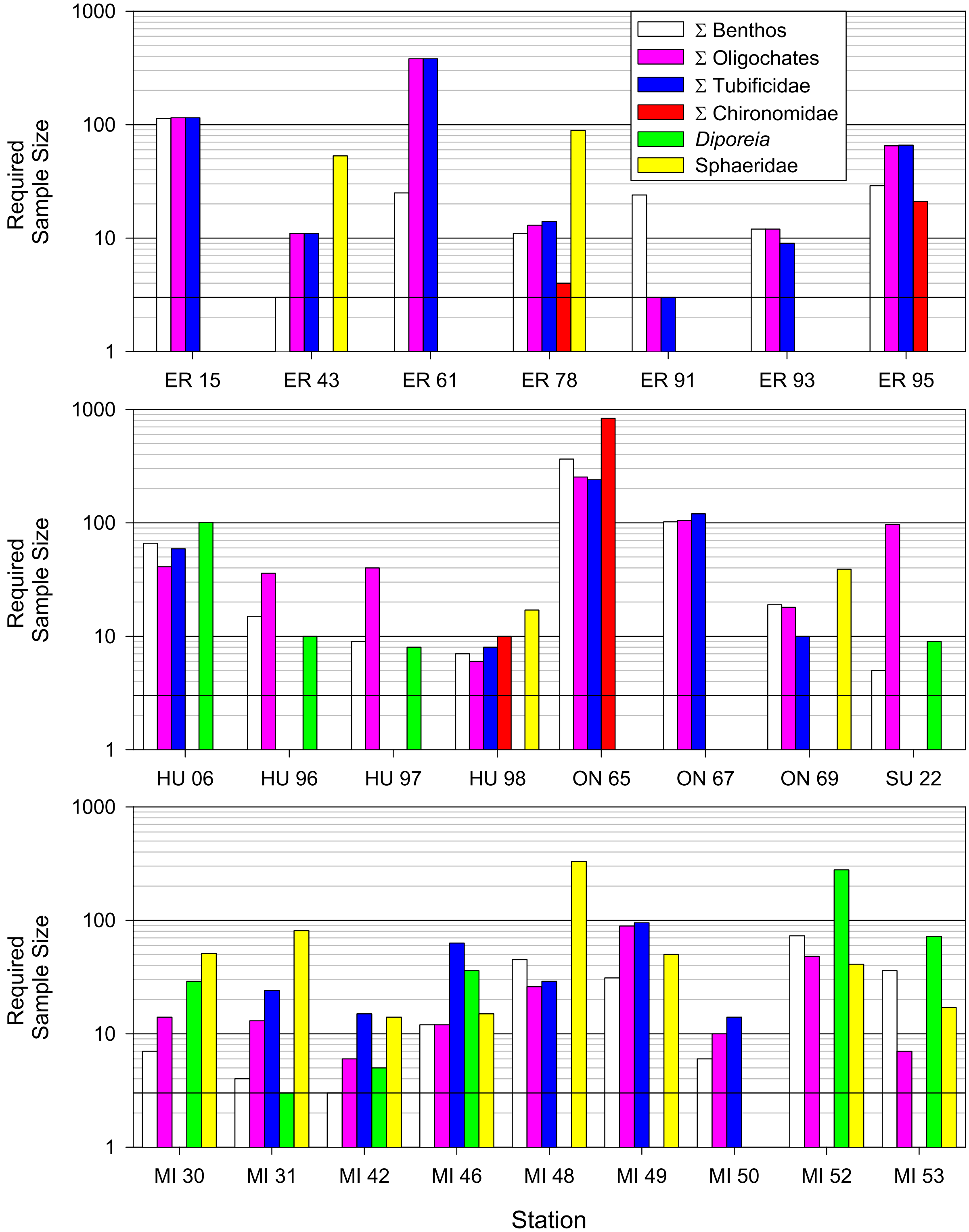

Figure 10. Sample sizes required at nearshore sites to detect a 20% change in total benthos, total oligochaete, total Tubificidae, total Chironomidae, *Diporeia* and total Sphaeridae areal densities when comparing 5 years, assuming $\alpha = 0.10$ and $\beta = 1 - 0.8$. Variance estimates and initial densities are based on 1999 values. Reference line represents n = 3.

For this analysis, only those sites for which all five years of data were available were used. Within site differences in densities between any two years were assessed using a one way ANOVA with year as the factor. This corresponds to the approach adopted above, and is the interpretation of the DQO used throughout this report. In accordance with the DQO, a significance level of $\alpha = 0.10$ was used. In the case of *Diporeia*, what is of interest is specifically a [decreasing] trend over time, rather than any year to year differences. To test for this, the non-parametric Spearman rank correlation coefficient was calculated for each site. While data from most sites conformed to the assumptions of normality and homoscedasticity, and in cases where these were not met, deviations were relatively small, a non-parametric test was used for correlations to enable detection of any, rather than just linear, relationships between year and density. For consistency, a significance level of $\alpha = 0.10$ was used.

Both the maximum difference in areal densities between any two years, and the maximum difference between any two consecutive years, were calculated for each site. Maximum differences between any two years ranged from $108/m^2$ to $10{,}510/m^2$, with a median value of $1{,}399/m^2$ (Table 4). Since the DQO is formulated in terms of a percent, maximum percentage differences were also calculated, using a conservative approach in which the larger of the two numbers was used as the denominator. Thus the equation used was:

$$\frac{MaxAnnual - MinAnnual}{MaxAnnual}$$

Maximum percentage differences ranged from 43% to 100% (i.e., where the minimum annual value = 0). The median percent difference was 72%, and 80% of values were greater than 55%. Therefore, in all cases an interannual difference substantially higher than the target detection limit of the DQO was seen. Since an alternate interpretation of the GLNPO DQO requires the detection of a 20% change between values in consecutive years, maximum differences between consecutive years were also computed for all sites. In this case, maximum differences ranged from $108/m^2$ to $5{,}487/m^2$, with a median difference of $1{,}316/m^2$. These corresponded to percentage differences ranging from 29% to 100%, with a median of 64%. 80% of sites had a maximum difference between consecutive years of 49% or more. Thus all sites experienced interannual changes in *Diporeia* density greater than the 20% targeted in the DQO, even when these differences were limited to those occurring in consecutive years.

Using the criterion for significance stated in the DQO ($\alpha = 0.10$), 26 of the 30 sites examined were found to have statistically significant interannual differences in *Diporeia* densities, when assessed using a one way ANOVA (Table 4). Since interannual differences can be a consequence of natural fluctuation in population sizes, rather than an indication of a trend, which is presumably of greater inherent interest, Spearman rank sum correlation coefficients were also calculated. In this case, 15 of the 30 sites examined exhibited statistically significant increasing or decreasing trends with time, indicating that in 11 of the cases for which interannual differences were detected, these differences were not directional.

Table 4. Analysis of *Diporeia* areal densities at all sites for which five (1997 - 2001) years of data are available. Maximum differences (Δ) in density estimates between any two years, and between any two consecutive years, are shown as both absolute values (#/m²) and percentages. Results of one way analysis of variance, with year as factor, and Spearman rank correlation analysis, are shown. Significant results, determined at $\alpha = 0.10$, are shown in bold.

Site	Maximum Δ (any years)		Maximum Δ (consecutive years)		ANOVA		Spearman Corr.	
					F	*P*	r_s	*P*
HU 06	3,996	100%	2,428	61%	**16.6**	**<0.001**	**-0.95**	**<0.001**
HU 32	3,060	66%	1,791	49%	**6.9**	**0.006**	**1.78**	**<0.001**
HU 38	1,358	73%	1,358	73%	**12.1**	**<0.001**	-0.34	0.209
HU 48	2,830	83%	1,740	75%	**6.7**	**0.007**	**-0.46**	**0.083**
HU 54	4,871	85%	2,014	70%	**29.8**	**<0.001**	**-0.68**	**0.005**
HU 61	1,759	48%	1,128	31%	**5.1**	**0.017**	**-0.74**	**0.001**
HU 93	2,435	79%	1,848	65%	**88.0**	**<0.001**	**-0.86**	**<0.001**
HU 95	2,282	71%	2,263	70%	**15.0**	**<0.001**	-0.31	0.257
HU 96	3,295	86%	1,402	53%	**5.3**	**0.015**	**-0.81**	**<0.001**
HU 97	1,574	49%	931	29%	**11.8**	**<0.001**	-0.35	0.194
MI 11	1,396	63%	1,396	63%	**2.7**	**0.096**	-0.29	0.293
MI 18	3,811	84%	2,855	63%	**6.2**	**0.009**	**-0.69**	**0.004**
MI 27	4,512	71%	4,512	71%	**9.1**	**0.002**	-0.13	0.629
MI 41	5,481	95%	4,640	80%	**18.3**	**<0.001**	**-0.74**	**0.003**
MI 46	1,294	43%	1,294	43%	**3.7**	**0.042**	**0.53**	**0.061**
MI 48	918	100%	918	100%	**16.0**	**<0.001**	0.17	0.578
MI 50	1,434	100%	1,338	93%	**6.7**	**0.007**	**-0.82**	**<0.001**
MI 52	10,510	100%	5,487	78%	**84.8**	**<0.001**	**-0.90**	**<0.001**
ON 25	778	57%	720	53%	1.8	0.214	-0.40	0.134
ON 41	1,377	71%	1,377	71%	**16.1**	**<0.001**	-0.18	0.514
ON 60	1,402	55%	1,134	45%	**4.4**	**0.025**	**-0.81**	**<0.001**
SU 01	535	48%	351	31%	**4.9**	**0.019**	0.05	0.842
SU 10	255	63%	249	62%	**3.5**	**0.049**	-0.29	0.287
SU 13	185	81%	178	78%	2.1	0.161	-0.41	0.124
SU 15	433	68%	433	68%	**3.7**	**0.043**	-0.22	0.410
SU 16	153	80%	121	63%	**3.2**	**0.061**	**-0.49**	**0.059**
SU 17	108	85%	108	85%	2.1	0.152	-0.22	0.426
SU 19	223	81%	204	74%	**6.2**	**0.009**	**-0.59**	**0.020**
SU 21	166	49%	166	49%	1.7	0.23	-0.20	0.457
SU 22	1,211	54%	1,052	47%	**16.9**	**<0.001**	-0.16	0.566

8 Discussion

In almost all cases examined, the current level of sampling effort in the benthos program is inadequate to satisfy GLNPO's DQO, as presently formulated. Even when just considering the limited number of variables used in this report, the great majority of sites would require over a dozen replicate Ponar samples to meet the current DQO, and over half would require more than three dozen replicates. This degree of sampling effort is clearly infeasible in the content of the GLNPO monitoring program, and it is in fact unclear the extent to which it is required to make educated ecological management decisions using the benthos data.

A number of ambiguities exist in the current DQO, some of which are specific to its application to the biological program, and some of which are more general. As noted in the introduction, the current DQO does not make entirely clear what should be compared when assessing changes in the magnitude of a variable. For the most part in this report it was assumed that what the DQO required detection of was a change between any two years of data. However, as was seen with *Diporeia*, a 20% fluctuation in density between two years was the rule, rather than the exception, and while this was due in part to actual basin wide declines seen in populations of this organism in recent years, even those sites which did not exhibit a trend (i.e., for which r_s was not significant) showed differences between years that were several times greater than 20%. This suggests that the normal range of natural fluctuations in biological populations can be considerably higher than the target difference specified in the DQO, and therefore that the DQO is probably unnecessarily stringent for biological data. In spite of not meeting DQO requirements with regard to being able to detect changes in *Diporeia* populations, the current sampling program nonetheless found statistically significant differences in annual *Diporeia* densities at nearly every site for which data were available. As suggested above, in at least some cases these differences were probably just due to natural fluctuations, and would not necessarily be indicative of an overall change in the ecological character of the benthos.

More restrictive interpretations of the DQO would include the ability to detect a 20% change between consecutive years, and being able to detect a trend in a variable. In the former interpretation, as seen with the *Diporeia* data, a 20% change might still be too stringent a requirement for benthos data; nearly every site examined exhibited a difference between consecutive years at least twice this great. In the case of the latter interpretation, unless a linear trend were assumed (an assumption for which there would be no *a priori* support), it would be difficult, and likely counter productive, to set a specific criterion for the magnitude of the trend to be detected. Presumably the detection of any trend would be of interest. Instead, a DQO might best be stated in terms of detecting a relationship between a variable and time (e.g., year) of a given strength.

The main deficiency in the current DQO, as it relates specifically to biological data (including the benthos data), is that the variable of interest is not specified. When assessing changes in the chemistry of a lake, determining the variable of interest is usually relatively straightforward (though not necessarily; e.g., instances involving detection of a large number of congeners of an organic pollutant). Such is not the case with biological data, however, which is multivariate. As pointed out in the introduction, potential variables include each individual species (of which there might be dozens, or even hundreds), groupings of species under broader taxonomic categories, community-level metrics such as species richness or diversity, and specific indices such as Goodnight-Whitley (Goodnight and Whitley, 1960) or Milbrink's (Milbrink, 1983). Between 1997 and 1999, a total of 86 taxonomic entities were recorded from the benthos samples collected by GLNPO, each one constituting a potential response variable to which the DQO could be applied. In many cases these species appeared in relatively low numbers at the sites at which they were found, which makes a given percent difference more difficult to discern, and also more difficult to interpret. This report limited the taxonomic groups under consideration to only those few that contained relatively high densities of individuals. Even with this restriction, sample sizes required to meet the DQO were unrealistically high. If the number of replicates required at each site were determined by that needed to meet the DQO requirement for *all* species encountered, the resulting sampling effort would undoubtedly become not only infeasible, but physically impossible.

It is also unclear how such changes, if detected, would be interpreted; e.g., what ecological meaning a 20% increase or decrease in *Stempellinella* would have. In some cases, changes in the biological populations of certain species are of inherent interest. One such case is *Diporeia*, as noted above. Large variations in the population sizes of dominant organisms might also be of interest, though the ecological meaning of such variations might not be readily apparent. Currently, the DQO does not provide guidance on which species or taxonomic groups are of interest, and therefore on which variables should be subject to the target contained in the DQO. Ideally, the benthos data should also be assessed in a way that enables detection of changes in the entire benthic community. This would require adopting a multivariate approach towards data analysis, and would probably require reformulation of the DQO to specify a probability of detecting a deviation from a 'baseline' community, rather than trying to detect a percent change in a single variable. This is because of the difficulty of reducing multivariate data to a single metric. As such, though, the definition of a baseline community, as well as what constitutes an ecologically significant deviation from such a community, would need to be defined. There is some precedent for such an approach in the Great Lakes (Reynoldson and Day, 1998); however further consideration of this alternative is beyond the scope of the present report. Ultimately, however, a more detailed statement of the DQO target with regard to the benthos data will depend on an explicit understanding of what the data will be used for, and how they will be interpreted.

9 References

Brinkhurst, R.O. 1967. The distribution of aquatic oligochaetes in Saginaw Bay, Lake Huron. *Limnol. Oceanogr.* 12:137-143.

Dermott, R. and K. Corning. 1988. Seasonal ingestion rates of *Pontoporeia hoyi* (Amphipoda) in Lake Ontario. *Can. J. Fish. Aquat. Sci.* 45:1886-1895.

Gardner, W.S., M.A. Quigley, G.L. Fahnenstiel, D. Scavia and W.A. Frez. 1990. *Pontoporeia hoyi* - a direct trophic link between spring diatoms and fish in Lake Michigan. In *Large Lakes: Ecological Structure and Function*, eds. M.M. Tilzer and C. Serruya, pp. 632-644. Springer-Verlag, New York, NY.

Goodnight, C.I. and T.S Whitley. 1960. Oligochaetes as indicators of pollution. *Proc. 15th Industr. Waste Conf. Purdue Univ. Ext. Eng.* 106:139-142.

International Joint Commission. 1978. *Great Lakes Water Quality Agreement of 1978.* Ottawa: International Joint Commission.

Milbrink, G. 1983. An improved environmental index based on the relative abundance of oligochaete species. *Hydrobiologia* 102:89-97.

Reynoldson, T.B. and K.E. Day. 1998. *Biological Guidelines for the Assessment of Sediment Quality in the Laurentian Great Lakes.* National Water Res. Inst. Rpt. No. 98-232, Environment Canada, Burlington Ont.