REGIONAL COPPER-NICKEL STUDY

BIOLOGICAL MONITORING

OF AQUATIC ECOSYSTEMS

Minnesota Environmental Quality Board
November, 1978

REGIONAL COPPER-NICKEL STUDY
BIOLOGICAL MONITORING
OF AQUATIC ECOSYSTEMS

Minnesota Environmental Quality Board
Judith E. Baxter
November, 1978

Abstract

Biological monitoring attempts first, to determine the status of natural biological communities and second, to detect changes from normal structure and function. It has theoretical advantages over physical and chemical monitoring in detecting the impact of mixtures, spills, fluctuating conditions, and other complex stresses. However, high natural variability makes it necessary to conduct well-designed intensive sampling programs both before and during an operation if any but very large (catastrophic) changes are to be detected. Even when changes are detected, the problem of determining what stress or combination of stresses are responsible remains. Experiments are needed to provide firm evidence in support of a postulated causal relationship.

In designing monitoring programs, the acceptable risks of Type I and Type II errors must be considered. If sampling programs are not intensive enough, the risk of failing to detect a significant impact of a given magnitude may often be as high as 50%. Such a high probability of Type II error would negate the value of a monitoring program. The magnitude of change to be detected, time and resources available, and statistical design used must also be considered in planning.

Monitoring programs may focus on single sites, paired sites, or multiple sites. The paired-site approach appears most promising because it reduces the effect of fluctuating environmental conditions and allows multiple sampling dates to be used without reduction to annual means.

Data from the Regional Copper-Nickel Study are useful in providing a gross baseline against which large (catastrophic) changes can be detected. They can also be a source of variance estimates used to determine numbers of

sites, sampling frequency and duration, and numbers of replicate samples for a planned program.

Recommendations on use of biological monitoring, taxonomic analyses, numbers of samples, and sampling methods for the Regional Copper-Nickel Study Area are made.

TABLE OF CONTENTS

INTRODUCTION TO THE REGIONAL COPPER-NICKEL STUDY

The Regional Copper-Nickel Environmental Impact Study is a comprehensive examination of the potential cumulative environmental, social, and economic impacts of copper-nickel mineral development in northeastern Minnesota. This study is being conducted for the Minnesota Legislature and state Executive Branch agencies, under the direction of the Minnesota Environmental Quality Board (MEQB) and with the funding, review, and concurrence of the Legislative Commission on Minnesota Resources.

A region along the surface contact of the Duluth Complex in St. Louis and Lake counties in northeastern Minnesota contains a major domestic resource of copper-nickel sulfide mineralization. This region has been explored by several mineral resource development companies for more than twenty years, and recently two firms, AMAX and International Nickel Company, have considered commercial operations. These exploration and mine planning activities indicate the potential establishment of a new mining and processing industry in Minnesota. In addition, these activities indicate the need for a comprehensive environmental, social, and economic analysis by the state in order to consider the cumulative regional implications of this new industry and to provide adequate information for future state policy review and development. In January, 1976, the MEQB organized and initiated the Regional Copper-Nickel Study.

The major objectives of the Regional Copper-Nickel Study are: 1) to characterize the region in its pre-copper-nickel development state; 2) to identify and describe the probable technologies which may be used to exploit the mineral resource and to convert it into salable commodities; 3) to identify and assess the impacts of primary copper-nickel development and secondary regional growth; 4) to conceptualize alternative degrees of regional copper-nickel development; and 5) to assess the cumulative environmental, social, and economic impacts of such hypothetical developments. The Regional Study is a scientific information gathering and analysis effort and will not present subjective social judgements on whether, where, when, or how copper-nickel development should or should not proceed. In addition, the Study will not make or propose state policy pertaining to copper-nickel development.

The Minnesota Environmental Quality Board is a state agency responsible for the implementation of the Minnesota Environmental Policy Act and promotes cooperation between state agencies on environmental matters. The Regional Copper-Nickel Study is an ad hoc effort of the MEQB and future regulatory and site specific environmental impact studies will most likely be the responsibility of the Minnesota Department of Natural Resources and the Minnesota Pollution Control Agency.

INTRODUCTION

As a part of the Regional Copper-Nickel Study (Study),aquatic communities of a large area (2000 square miles) were sampled both qualitatively and quantitatively over a period of two years. These data were collected primarily to provide a regional characterization of these communities, which could be used to predict changes due to various stresses in these communities. Such a regional characterization gives an indication of what species and groups of species would be expected in different types of streams and lakes of the area. The regional characterization of these systems is summarized in the Aquatic Biology characterization report (2nd level) and is presented in detail, by taxonomic groups, in other Study reports (Regional Copper-Nickel Study 1978 a; b; c; d; e; f; g).

However, these data were also collected to provide a basis for the design of biological monitoring programs for potential copper-nickel mining developments. It is the purpose of this report to discuss factors which must be considered in designing biological monitoring programs and several designs which may be used. This report also discusses the ways in which data collected by the Study may be used in designing site-specific studies and the extent to which they provide a "baseline". Some of the data are used to illustrate part of the process of designing a biological monitoring program, and some general recommendations for site-specific monitoring in the Study Area are made.

CONCEPTS OF BIOLOGICAL MONITORING

In general, biological monitoring attempts first, to determine the status of natural biological communities and second, to detect changes from normal structure and function so that stresses can be reduced, removed, or avoided in future development. Implicit in the concept of detecting changes from

the "normal" is the idea of establishing the "baseline" condition (i.e. the "normal" against which changes can be detected). Since natural populations normally fluctuate, the problem is to detect "unnatural" changes against a background of natural ones. Mount (1976) suggests that such monitoring may also be utilized to compare the impact of different discharges, to develop priorities for abatement, and to verify the effectiveness of standards by verifying the effects of abatement procedures.

It should be noted that the term biological monitoring has also been used to refer to in-plant or in-receiving system monitoring units which monitor the responses of test organisms to effluents as they are discharged . (Cairns 1976, Morgan 1976, Poels 1976). This type of monitoring can more appropriately be called biological "effluent monitoring". This approach has been suggested because changes in environmental parameters (pH, temperature, etc.) can modify the response of organisms to specific toxicants (Cairns 1976). This monitoring of effluents provides a biological measure of the effects of the real mixture of toxicants in an effluent. It assumes that the test organisms adequately represent all the organisms of the exposed biological community. It may in fact provide the best warning of problems. However, the subject of this report is monitoring of natural biological communities.

Traditionally, monitoring associated with concern about water pollution has emphasized physical and chemical parameters because these are easier to define and evaluate and less expensive to measure than many biological parameters. There are several advantages of adding biological monitoring to physical-chemical monitoring. For instance, Gaufin (1973) stated that

> "Since chemical studies give information on physical-chemical conditions only at the time of sampling, and pollution surveys frequently cannot be made during the period of the most critical conditions, there is need for additional methods that can be used throughout the year

> for determining the extent and severity of brief critical or limiting environmental factors. The qualitative and quantitative composition of an aquatic population is determined by recurring critical conditions, even though of short duration, as well as the more stable or long-term environmental factors. Therefore, the complex of organisms which develops in a given area is, in turn, indicative of environmental conditions which have occurred during its development. Organisms having life histories of a year or more will thus serve to indicate unfavorable or limiting conditions that have occurred several months previously. Because aquatic populations are a result of past environmental conditions, they serve as a means for determining such conditions in a stream. They are especially valuable because they can be used during fall, winter, or spring months, when flows may be large, dilution is a maximum, dissolved oxygen is near saturation, and visual evidence of pollution at a minimum, to delineate former septic areas or to indicate critical conditions of short duration."

Cairns et al. (1973) have presented a similar argument:

> "Perhaps the most important reason for doing biological monitoring is that aquatic organisms act as natural monitors. During a short-term exposure to water of poor quality, organisms that cannot tolerate the stress are destroyed and the aquatic community structure changes. Since aquatic organisms respond to their total environment, they provide a better assessment of environmental damage than do the handful of chemical or physical parameters (dissolved oxygen, temperature, conductivity, pH, turbidity, etc.) that can now be continuously monitored effectively. It is important to recognize that biological monitoring does not replace chemical and physical monitoring. They all provide converging lines of information that supplement each other but are not mutually exclusive. A biological monitoring program is essential in determining the synergistic or antagonistic interactions of waste discharges and the receiving system."

However, those who point out the advantages of biological monitoring in detecting effects of multiple stresses often do not make clear the problems involved. Biological monitoring programs are usually undertaken to detect the impact of some human activity. Because the detection of impacts must occur in a natural community, in relation to an actual development, the controls usually employed in scientific experiments are difficult or impossible to use. Natural variability and determination of cause-and-effect relationships then become major problems.

Natural variability is a problem because large fluctuations in population size occur in natural ecosystems even without man's influence. Some of the factors that have been implicated are external to the biological systems, such as weather, temperature, photo-period, flow, pH, while others are internal, such as life cycle, food supply, predators, and competition. The interactions between most of these factors are poorly understood. Natural variability can mask changes due to human interventions until they become large or catastrophic for the populations involved. Hirsh (1976) suggests that to protect our environment, it is preferable to detect the "subtle, sub-acute, non-catastrophic changes" which presumably precede catastrophe, in order to avoid ecological disasters. To detect non-catastrophic changes in the face of natural variability, a well-designed, fairly extensive sampling program is necessary. Necessary considerations in designing sampling programs are discussed later in this paper.

Determination of cause-and-effect relationships is a problem because human activities rarely produce only one, simple change in environmental factors. Hirsh (1976) has discussed the problem of attributing a change to a specific environmental stress and points out that this is feasible in "cases where the source and nature of the disturbance are relatively discrete and the effects are relatively well understood (e.g. site specific studies of the impact of industrial pollutants on streams)". When there are multiple sources of different pollutants and pathways are complex, it is difficult to identify the stresses responsible for a change. Without identifying the cause of changes, it is impossible to reduce or eliminate the stress.

For example, consider a shift in dominant organisms after the introduction of a pollutant into a stream. It is first necessary to determine if the tolerances of the various species for this substance(s) are consistent with

the shifts in their abundance. If they are consistent it may be reasonable to attribute the observed changes to the presence of the pollutant.

However, if there is more than one pollutant, in varying concentrations over time, consistency is difficult to establish. Studies of heavy metals pollution, where effluents tend to contain mixtures of metals, have rarely identified the specific components responsible for changes. Synergistic effects may be suspected, but are difficult to confirm in natural systems. Similarly, Hirsh (1976) cites another example of longterm trends in marine zooplankton populations which appear related to man's activities, but whose specific cause is unknown. Many environmental changes coincide in time with the observed zooplankton changes, and it is difficult to isolate the causal factors.

It is clear that the detection of changes through a monitoring program can create the need for further studies, but may not answer the question of what factor is responsible for the changes or suggest how to reduce such impacts. In general, experimental tests of hypotheses are necessary to provide firm evidence in support of postulated cause and effect relationships. These are difficult to perform without modifying the natural system of interest. Ecological experiments such as the clearcutting studies at Hubbard Brook (Likens et al. 1969) are only possible in a limited number of cases. In these cases, realistic treatments simulating industrial development are applied to sites previously monitored to establish their relationship to central sites. They are similar to the paired site studies discussed below.

Statistical concepts

The problem of planning a biological monitoring program is one of designing a test to detect abnormal change in a system with high natural variability,

where the mechanisms of change are poorly understood. Because some change (i.e. variability) is considered normal, it is necessary to use statistical testing procedures in order to determine the likelihood that an observed change is significantly different from the norm (baseline). In order to apply statistical procedures, the investigator must specify both the magnitude of change that is to be detected, and the probability of error that is acceptable in the determination.

The magnitude of change considered significant has rarely been explicitly stated in environmental monitoring programs. However, it is very important because it directly affects the design of the sampling program. Because the significance of a change can only be established in relation to the natural variability of the system, the number of samples and the length of the sampling program limit the amount of change that can be determined to be significant statistically. To detect "subtle" (small) impacts on the systems it is necessary to conduct long and intensive sampling programs both prior to and during "pollution episodes". On the other hand, it may be possible to detect "catastrophic" (very large) changes with fewer samples over a shorter time period. Once the magnitude of impact to be detected has been determined, the design of the sampling program must take into account the possibility of an erroneous conclusion and an acceptable level of risk must be determined, Constraints of time and money must also be considered. The amount of risk that is acceptable depends on one's point of view. Eberhardt (1978) explains these considerations very well.

> "Consider the likely points of view as to environmental impact of new construction by the executives of an industry and of a regulatory agency, respectively. Suppose that both parties agree to a survey method that has well-known statistical characteristics, and that they further agree to make certain modifications on factory operations if the field survey shows a specified degree of change has taken place. All that remains is to decide how large a sample should be taken. But that depends on the amount of protection each party requires against errors damaging to its own best interests:

1) The manufacturer would rather not have the survey results indicate a significant change when the agreed on degree of change really did not take place (Type I error). Just how firm the industry representatives may be on this point is likely to depend on costs of corrective or alternative measures. If the changes are minor, perhaps they will agree that a 10 percent rate of such "false positive" results is acceptable. If costs of the changes are quite expensive, they well may want to insist on error rates of 1 percent or less.

2) On the other hand, the staff of the regulatory agency would rather not fail to recognize a significant impact when one takes place (Type II error). If very small samples are used, the results almost always will come out not significantly different. Hence, the agency representatives may be guided by rules that indicate that there will be an 80 percent chance of being sure to detect a real difference (of the magnitude agreed on) when the environmental impact is not of minor consequence. On the other hand, if very substantial damage to an important resource possibly is involved, they well may want to insist on something like 99 percent assurance. All too often, whether by default or lack of understanding, the real rate is about 50 percent, which can be likened to settling the issue by flipping a coin and doing no field study whatsoever.

The first kind of error (detecting an impact that is not real) described above is reasonably well known, probably because it must be specified for virtually any kind of statistical test. Most biologists seem to have settled on about a 5 percent rate for this kind of error. As with any percentage or proportion, there are two ways of specification--a 5 percent chance of error or a 95 percent 'assurance" that the statement is correct."

The probability of Type II error (failing to detect a change of the agreed-upon magnitude when it occurs) is rarely discussed in biological studies. The probability of Type I error (α) is usually chosen by the investigator and is fixed for all sample sizes. The probability of a Type II error ($1-\beta$) decreases as sample size increases, for fixed α. Thus, ideally, one should take as large a sample as possible, as larger samples decrease the probability of Type II error. Then power curves can be inspected to determine the tradeoff between Type II errors and detectible differences. In some instances, it may not be possible to take enough samples to reduce α and β to acceptable levels. In these instances, the decision must be made

whether to accept detection of larger changes only, or to make other changes in objectives. Without consideration of these problems, monitoring programs cannot be designed to fulfill their objectives. The actual probabilities of error achieved in a monitoring program depend on the actual variances and differences observed, and may be quite different from those on which earlier calculations were based.

DESIGNS FOR MONITORING PROGRAMS

In designing a monitoring program, it is necessary to acknowledge that no two areas and therefore, no two populations are identical; by taking enough measurements, this can always be detected statistically, although many attributes are similar. On the other hand, natural populations normally change from year to year and from season to season. The problem is to detect "unnatural" changes against a background of natural variations in time and space. "Unnatural" changes are changes in the parameters of the distribution. Extreme values will occur occasionally by the laws of probability.

There are three basic approaches one can taken in setting up a monitoring program under these conditions:

1) consider one site before and after some operation begins;

2) consider pairs of sites (control and treatment sites); and

3) analyze a series of sites subjected to a gradient of stress, using a regression approach. Each of these approaches requires a different sampling program.

Single Sites

It is intuitively appealing to consider a single site before and after some environmental change. The natural variability in the parameter of interest

would be determined over a period of time in which human disturbance is limited. Then, once some development occurs, comparisons would be made between data collected before and after the development.

The major problem with this approach is that of separating changes caused by seasonal factors and/or normal year-to-year differences from changes caused by the development being studied. Seasonal patterns of physical and chemical parameters change between years. Because these parameters affect the life cycle of aquatic organisms, seasonal changes in aquatic organisms vary from year to year. For example, the spring emergence of the mayfly Leptophlebia will be influenced by the timing of spring runoff and the pattern of water temperature, which vary considerably from year to year. Therefore, data collected on the same date in two years could indicate a great change in population size of Leptophlebia when in fact it was only the time of spring emergence which had changed. On the other hand, the peak density of Leptophlebia will vary from year to year and be influenced by water temperature, pH, and a host of other factors.

Because of these seasonal patterns, if sampling is done at one point in the year, then that point must be timed precisely, every year, to occur at comparable points in seasonal cycles. Such comparable points are difficult to determine. Alternatively, sampling can be spaced throughout the life cycle, and the baseline can be examined at the level of annual means of sample values. Then variations in timing of the samples are assumed to be random, and the mean of the sample values is used to represent the year. In both approaches, it is necessary to establish a long record to determine the natural limits of fluctuations and establish the normal between-years variance in parameters. The number of years required depends on the actual variance and the magnitude of changes that the program is designed to detect.

In some cases, three or four years, the absolute minimum time required to calculate a variance, would be sufficient. In other cases, one hundred years might be inadequate to detect small but ecologically important changes.

An example from Study data may clarify these points. Figure 1a shows seasonal changes in mean diversity of diatoms at station P-1 on the Partridge River. The mean diversity was calculated as the mean of the Shannon-Wiener diversity Index ($H' = \Sigma P_i \log_2 P_i$) for each of two to four glass-slide artificial substrate samplers. Annual means based on samples from four dates are shown.

The differences between the mean diversity at one date in 1976 and (roughly) the same date in 1977 (Figure 1a) are much greater than the differences in the annual means (Figure 1d). This is consistent with Figure 1c, which shows that the seasonal pattern of discharge at this site was quite different in 1976 and 1977. It is clear that large differences in yearly samples at one point in time are normal. Only two years of data makes generalization risky, but it appears that variability in annual means is smaller, as would be expected. Programs based on only one sample per year would thus require many more years of sampling to detect changes than programs based on annual means.

The 95% confidence intervals for mean diversity of replicate slides (Figure 1b) suggest another problem. Values on individual slides are frequently quite different from the mean values. Thus we would not expect small changes in mean values to be statistically significant unless many replicates were examined. Since seasonal changes are clearly large, it is preferable to sample several times a year rather than to sample very intensively at one point in time.

Paired Sites

The second approach is to employ what Eberhardt (1976) has called a "pseudo-design," involving paired sites. He summarized the concept as follows:

> "...two areas that are not widely separated in space are ordinarily subject to much the same factors, have populations with about the same genetic make-up, and generally (two such) populations can be expected to follow much the same trend over time - apart from human intervention. These premises lead to the conclusion that about the best we can do currently with the 'single site' (treatment) situation is to require that there be one or more control sites, and that a'baseline' (pre-construction, pre-operational, etc.) period be used to establish the ratio of population density in the prospective 'impacted' site to that on the control site(s)."

In practice, several pairs of sites provide replication for such a design. For simplicity, Eberhardt only discussed population density as a parameter, but he notes the need for analysis of what other parameters might be treated in a similar fashion. Diversity, or relative abundance of important species or functional groups (Regional Copper-Nickel Study 1978 e) could also be used, though not necessarily compared by ratios. The basic concept is that the relationship between two stations in a pair is established during pre-operational study and should continue in the operational period, thus any changes in the relationship are attributed to the operation being studied. Eberhardt calls such a study a "pseudodesign," in contrast to a true experimental design, because the analyses commonly used for true experimental designs rest on the assumption of random assignment of experimental treatments and control conditions to a considerable number of test plots, animals, etc. and can thus separate out variability not connected to the factor of interest. This is generally not possible in environmental monitoring situations because there is no certainty as to which site(s) will be subjected to stress.

Ideally, biological sampling should be preceded by physical and chemical sampling to establish the range of values characteristic of the region. Then control-treatment ("impacted") pairs of sites can be selected on the basis of physical and chemical similarity. Eberhardt (1976) notes that unexpected things may happen to control sites, so that it is advisable to include extra control sites in a sampling program.

The major problem lies in choosing control stations sufficiently far from the operation to be beyond its influence, but close enough to the "impacted" stations to be expected to be subject to the same natural environmental changes. Stations upstream and downstream from a source in streams, with similar physical characteristics (currents, substrates, canopy cover) are good possibilities. Sites near and far from the path of a discharge in a lake are also possibilities. Therefore, it is absolutely necessary to define the physical paths of the potential stress in selecting sites. The monitoring can then be designed to detect the effects of a particular stress.

The advantage of the paired-site approach is that the pairing of sites reduces, in theory, the effects of fluctuating environmental factors in the analysis, since they are presumed to affect both sites similarly. This reduces the error variance in the analysis of variance, and reduces the number of sample dates and pairs of sites needed to detect a given magnitude of change. Because the paired sites presumably respond in similar ways to external factors varying over time, all sampling dates can be used in the analysis, without reducing them to annual means, thus the time needed to establish a baseline is reduced. However, the length and intensity of the sampling program is still crucial.

McKenzie et al. (1977) employed the paired-site approach in analyzing many sets of data collected in monitoring programs at nuclear power plants. They discuss in detail the many statistical decisions to be made in planning a program, and models and transformations that can be used. Their results, as summarized in Table 1, show that sampling programs of the magnitude of those currently being used at power plants can detect changes of from 20 to 50% of expected levels, with probability of Type I and Type II errors (α and β) controlled to be respectively 10 and 20%. They recommend use of an alpha level of 10% rather than the more commonly used 5% in order to achieve acceptable power ($1-\beta$) at sample sizes currently in use and considered feasible. They found that paired-site designs were much more efficient than the commonly used unpaired analysis of variance methods. Thus, if the variability in the ecosystem being studied is similar to conditions in the studies examined by Mckenzie et al. (1977) similar magnitudes of change could be detected with similar sampling programs.

Multiple Sites

In some situations, a regression analysis, somewhat similar to the paired-site approach, can be used. The requirements are that a readily measured gradient of "insult" exists and persists in time and "that effects are sought in measurements on sessile or sedentary elements of the biota " (Eberhardt 1976). For example, in streams, attached algae are more suitable than fish for such studies. This can be seen as an extension of the paired-site approach where the low end of the gradient acts as a control site and the other sites have various levels of "treatment." Several studies of the effects of the smelter at Sudbury, Ontario on lakes at different distances from the stack (Stokes et al. 1973, Gorham and Gordon 1960) have taken this approach. If a clear gradient of effect on the biota is demonstrated, it may not be

necessary to demonstrate that the sites were similar before the operation began, but if the pattern of effects does not follow a clear gradient, then the paired site approach is probably more useful. Modeling of these effects may also be combined with the paired-site models using analysis of covariance techniques (Cochran and Cox 1957).

PLANNING A MONITORING PROGRAM

The paired-site (treatment and control) approach appears to be the most promising way to utilize aquatic biology monitoring data. However, the statistical problems involved are complex, and many pitfalls await investigators who fail to invest considerable time in consideration of their choices before sampling begins. Biologists should work intensively with statisticians before designing such a program, particularly with respect to defining the levels of change they need to detect and the risks they will accept. Useful references are Eberhardt (1976; 1978) Hirsh (1976) and McKenzie et al. (1977).

The following questions must be answered while developing a program:

1) What organisms should be monitored?; and

2) How many samples should be taken?

Choice of Organisms

The defects of the shotgun, "sample everything" approach have been discussed often (Hirsh 1976, Eberhardt 1976, 1978; Sharma et al. 1975). This approach results in such a dilution of sampling effort for any particular group that only the grossest, catastrophic kinds of effects can ever be detected. Also, multiple comparisons can result in the probability of Type II

error being dramatically increased. Kaesler (1974) has shown that when information on many groups of organisms is collected at the same sites, there is redundancy in the data. In other words, data in one subgroup may carry much of the information in the whole set. McKenzie et al. (1977) point out that although most impact studies have identified organisms to the lowest taxonomic level possible, data analyses usually focus only on many fewer, higher level categories. Thus, costs could be reduced by taxonomic analyses appropriate to the level of data analyses. Mckenzie et al. recommended classification into general taxonomic categories (e.g. phytoplankton into diatoms, blue greens and greens) and retention of samples for further analysis if effects are seen. Some authors (Herricks and Shanholtz 1976, Kaesler et al. 1971) have used more detailed data with similarity coefficients and cluster analyses, but such analyses cannot provide estimates of what magnitude of change they can detect. There is little theory to indicate these techniques can detect small changes.

There is often great concern about impacts of developments on fish populations. Programs for monitoring fish populations face particular problems which have been summarized by Mckenzie et al. (1977). Van Winkle (1977) has enumerated four unavoidable limitations imposed by both the fish populations and the decision making process that contribute to the inconclusive nature of fisheries impact assessment. These four limitations are:

1) The existence of compensatory mechanisms and our inability to quantify them;

2) Large, uncontrollable and unpredictable natural variations in reproductive successs and stock size, including the general lack of understanding of the stock-recruitment relationships;

3) The time frame for decision-making is generally less than the generation and response time of the fish populations of interest; and

4) Uncontrollable and unquantified effects from other sources of impact.

Another limitation is the migratory behavior or open system aspects of most of the populations of interest, i.e., the sport or commercial species. Realistically, the relationship between the impact of nuclear power generation and the fish populations is an area which requires further research and quantification.

The choice of what to monitor is not easy, but it should be noted that monitoring too many things may have the same effect as not monitoring at all. As Eberhart (1978) has pointed out, a sampling program without careful statistical planning may often have only a 50% chance of detecting an impact at a level considered significant. In environmental monitoring, it is important to detect change if it occurs. Therefore, a test with so low a power is probably not worth using. If monitoring is to fulfill its presumed function of detecting change if it occurs, sampling effort must be well designed and focused on a few groups of organisms considered important.

Important species or groups may be endangered species, groups performing key functions in the ecosystem, economically or aesthetically important groups, or groups known to be sensitive indicators of changes of concern. The only generalization that applies is concentrating effort on whatever organisms or parameters are considered important in any way. Ideally, ecological importance should be a prime criterion, but unfortunately, the relative importance of many components of ecosystems is not well understood. (Eberhardt 1976).

Number of Samples

Five things must be specified to determine the required sample size:

1) The magnitude of effect to be detected;

2) The Type I error that is acceptable (α level);

3) The probability of Type II error that is acceptable (β-level);

4) The statistical design to be used to test for impacts; and

5) The time and resources available.

All of these factors must be considered in designing an effective monitoring program that will achieve its objectives.

In discussing the number of samples needed, some terminology should be clarified. In biological monitoring programs, one is generally interested in whether or not there is a change at a site, which corresponds to a sample plot in traditional experimental design. Thus one sample may be the data from one site at one sampling time. The use of replicate samples will reduce the estimate of variance between similar sites, in some designs, and will provide more reliable estimates of site means. However as Eberhardt (1978) notes,

> "one common mistake in the analysis of observational data is the assumption that variances based on subsampling of plots are suitable bases for comparing treatments. This usually is not a valid basis for a statistical test, because it is the plots-treated-alike variance that is relevant to a test of treatment differences. The essential issue is that no two areas can be expected to be identical. The whole idea of experimental design is based on the random assignment of treatments to plots and the use of variation between plots given the same treatment as a standard on which to judge whether real differences result from two or more treatments. Subsampling the treated plots may be necessary for various reasons, but for analysis it is the plots that are important."

Some confusion arises because it is generally easier to increase the number of sample replicates at a site than to increase the number of sites. However, to determine if differences between control and treatment sites are larger than the normal differences between any two untreated sites, data are needed from multiple control sites (or multiple dates from a control site). These data provide estimates of Eberhardt's plots-treated alike variance.

It should be noted that in sample size calculations using the paired-site approach, each sample point represents a ratio or difference of two values from the paired stations. If replicate samples are taken at each site, they enter the analysis only as mean values, and the variance associated with these replicates (which may be called subsample variance) is not used directly in the testing of hypotheses about changes in the ratio or differences between pairs of stations (Mckenzie et al. 1977, Eberhardt 1978). Thus the question of how many replicates to use must be decided separately from the sample size calculation which specifies the number of station pairs and sampling dates. Once a design is chosen and the magnitude of effect and possibilities of errors have been set, then sample sizes can be calculated, based on estimates of variance. These can come from a pilot study, or from existing data at other sites. The use of Study data to make such estimates is discussed below.

Unfortunately, it is these calculations of sample size needed to achieve desired objectives that are often omitted. Without them, biological monitoring programs will not meet the needs implicit in attempting to sample intensively. If a program is seen only as monitoring for catastrophic changes, such as the near extinction of major groups, many fewer samples are needed, as the presence or absence of fish, diatoms, etc. is easier to determine than estimates of their numbers, and normal changes in abundance.

Regional Copper-Nickel Study Data on Stream Biota

Effective biological monitoring requires careful planning, consultation with statisticians, and above all, clear determination of objectives so that effort can be focused. The data collected by the Study can be a valuable resource in such planning. However, it is impossible to anticipate exactly what data summaries or statistics would be useful in the future. The quantitative and

qualitative taxonomic data on stream invertebrates and periphyton are available on magnetic tape, together with programs for generating, mean values, confidence intervals for many parameters, similarity coefficients, tables of species abundances, and cluster analyses. In this report a small part of the data will be used to demonstrate how they might be used.

The Study collected extensive quantitative data on stream organisms in three ways:

1) Drift net samples of drifting benthic invertebrates;

2) Hester-Dendy samplers (artificial substrates) for benthic invertebrates; and

3) Glass slide artificial substrates for diatoms

Qualitative collections of invertebrates and periphyton were also made, but these data can only be used to indicate catastrophic changes. Stations were sampled with varying frequency, with larger streams (3rd and 4th order) as opposed to smaller tributaries (1st and 2nd order) sampled most intensively. Details of this program are discussed in Aquatic Biology Operation Manual (Regional Copper-Nickel Study 1977). Tables 1-6, Appendix A, show the number of samples of each type collected at each station for all sample periods.

Regional Copper-Nickel Study Data as a Baseline

The data collected by the Study can serve as baseline data in two very limited ways. First, if future developments occur near sites sampled by the Study, then some of these sites could serve as part of control/treatment pairs for these developments.

Secondly, the regional characterization provides a gross baseline against which near catastrophic changes could be detected. For instance, in the example used earlier (Figure 1) it appears that mean values for diatom diversity at P-1 based on 2-4 slides, are occasionally less than 1.0. Thus, even though 95% confidence limits for individual time periods included lower values for most dates in 1976, it appears that mean values that stay consistently below 1.0 would suggest an abnormal community. Examination of data from all the sites sampled shows that 89% of the diversity values are higher than 1.0, (Figure 2) which suggests 1.0 (in $\log_2$ units) as a "warning" level to prompt further investigations. However, it has been shown that diatom diversities in the Study Area tend to be low in summer (Regional Copper-Nickel Study 1978f). Thus non-summer samples should be examined before conclusions are drawn.

Normal Values for some Biological Parameters in the RCNS Study Area.

The range of mean values found by the Regional Copper-Nickel Study for diversity and number of taxa of diatoms on glass slides, number of taxa of invertebrates in drift nets and total of density of invertebrates in drift nets are shown in the frequency distributions presented in Figures 2-5. Mean values were all based on at least two sample replicates. These values are dependent both on the sampling techniques used and on the taxonomic level to which specimens were identified. Thus future comparisons should be made using similar procedures. Samples from the Erie Mining Study are discussed in a separate report (Regional Copper-Nickel Study 1978i).

The pattern (Figure 2) of diatom diversity values (Shannon-Wiener Index) has been discussed above as an example of the use of these data as a baseline. Figure 3 shows the distribution of values for mean number of taxa, based on

the same diatom analyses. It is clear that, as expected, the patterns are very similar. It appears that values of 6 to 30 are normal for the number of diatom taxa on glass slides.

The frequency distribution of the mean number of taxa found in drift samples is shown in Figure 4. When the values are separated into groups (not shown) comparing first and second order sites to higher order sites, a trend toward higher diversity at higher order sites is seen. In general, a very rough examination of the data suggests that fewer than 12 taxa in drift is unusual for higher order sites, and fewer than 6 is unusual for first and second order sites.

The frequency distribution of geometric means of total density of drift (Figure 5) is clearly skewed. Geometric means are given because the sampling distribution for total densities tends to be log normal; thus the data were subjected to a log transformation for calculation of means and confidence intervals. The mean of the logs is then the log of the geometric mean of the original densities. Stations E-1 and K-2 have the highest densities, and were also shown by water quality analyses high levels of nutrients. Examination of the graph suggests that geometric mean drift densities of greater than 3000/1000 m^3 flow would represent a change from normal conditions for most sites. Without more detailed analysis, it is not possible to suggest a value for abnormally low densities.

These graphs (Figures 2-5) are not based on random samples of sites of different orders with comparable sampling schedules, so they do not represent true distributions for these parameters in the region. However, they

do give some picture of the range of values found in the Study rea.

Use in Planning a Paired-site Monitoring Program

Planning a paired-site monitoring program requires the selection of a number of pairs of sites (or, preferably, groups of similar sites with one or more control sites matched to each potentially impacted site. It is important to determine sample size, which in this case is equivalent to determining how many site pairs will be sampled, how frequently, and for how many years prior to the operation. Once the parameter of interest (e.g. drift density) is selected, the level of change to be detected is set, and acceptable levels of error are chosen, then data from the Study can be used to estimate the sample size needed , if one assumes that some of the sites sampled by the Study are similar to the sites under consideration (see Appendix B). On the basis of the regional characterization of stream communities of the Study Area, it appears that stream order, water chemistry, and streamside vegetation should be the factors considered in selecting Study sites similar to other sites near planned operations (Regional Copper-Nickel Study 1978 e, f, g).

As used by McKenzie et al, (1977), the paired site approach is designed to test for the impact of the operation of some facility on a biological parameter using analysis of variance techniques. The models used would vary according to the particular situation involved. For instance, in lake sampling programs, pairs of sites might be selected to include a factor for depth, or placement in relation to a discharge, or substrate type. If such factors are controlled, the residual variance would be expected to decrease. The sites in a pair must be as similar as possible, but the site pairs do not have to be similar in their response to the stress. Differ-

ences in impact on site-pairs can be included in a model as an interaction term for a site-pair effect and a status effect. Given g pairs of sites sampled f times per year for h years per treatment, a typical model might be:

$$Y_{ijkl} = \mu + S_i + M_j + P_k + SM_{ij} + SP_{ik} + MP_{jk} + E_{ijkl}$$

where

Y_{ijkl} = The measured ratio, log ratio, or difference between values for a pair of sites for

status i (i = 1 preoperational)
(i = 2 post operational)
station pair k (k = 1,g)
year l (l = 1,h)
season j (j = 1,f)

μ = the overall mean value

S_i = Status effect (1 = preoperational, 2 = operational)

M_j = Season effect (j = i, f)

P_k = site-pair effect (k = i, g)

E_{ijl} = random error term, hopefully normally distributed, including all interaction terms not specified in the model.

SM, SP, MP are 1st order interaction terms.

The sites are sampled for h preoperational years, and for h years during the operation. Usually, McKenzie et al. (1977) used log (density at control site/density at treatment site) in studies of lake benthos, zooplankton, and phytoplankton. The model above is discussed for illustrative purposes only, and represents only one of many models which may be appropriate. The model used must be tailored to the situation being studied.

With such a model, the hypothesis of interest concerns the status effect,

or the significance of the S_i term. The hypothesis being tested is $S_1 = S_2$, or intuitively, that the Y value, or the log of the ratio of densities, was constant during preoperation and operational periods (i.e. no change occurred between preoperational and operational periods). Thus, the F test of interest tests the status effect against the residual variance.

Because the whole analysis is planned to test for the difference between two treatments (preoperation and operation), the number of samples needed (data points for each treatment) can be estimated based only on the size of difference to be detected, ($|d|$) and an estimate of the residual variance in the data (S) when month, site pair, and status effects have been removed (see Appendix B).

Data from the Study can indicate what kind of sampling program would be needed to detect particular levels of change. Consider the data on total density of drift per 1000 cubic meters of water at two 4th order stream sites: P-1 (Partridge River) and SL-1 (St. Louis River). These sites were grouped together by analyses of water chemistry data (Regional Copper-Nickel Study 1978 j) and have similar physical characteristics. Both receive mine drainage from taconite operations without leachate from copper-nickel deposits. Thus they could be expected to react in similar ways to environmental fluctuations. For each site on each date, a 95% confidence interval for the geometric mean density was constructed using a log transformation and the formula for confidence limits for a mean based on the t-distribution, when only s, the sample variance, is known. The transformation was used because higher densities tended to have larger sample variances. Thus the density was assumed to be lognormally distributed. These confidence intervals are shown in Figure 6 in terms of actual density

values. The confidence intervals for the two sites overlap on all dates, which supports the choice of the sites as similar.

In the paired site approach, one is interested in the relationship between the mean densities at the two sites. McKenzie et al. (1977) discuss alternative methods of analyzing the relationship between densities at paired sites. They suggest the use of logarithm of the ratio of the densities, on the hypothesis that the ratio of the densities is constant. Under this hypothesis, both sites respond to a change in the environment by the same proportional change in density. Figure 7 shows the log of the ratio of mean densities at sites P-1 and SL-1 over the time period sampled by the Regional Copper-Nickel Study. The ratio varies considerably over the two years sampled, and it appears that the relationship between the two sites was different in 1976 from 1977. The paired-site approach rests on the assumption that on the average, the relationship between the sites is constant with "normal" variability (i.e. random variation about a mean).

A test for the suitability of two sites as a control-treatment pair is proposed by McKenzie et al. (1977). They suggest that for two similar stations, the covariance of the variables being studied at the two sites should be positive. This tests whether the paired sites tend to vary in the same direction. Unfortunately, the covariance of the drift densities at SL1 and P1 is negative. Thus these do not appear to be as well matched as we would like, but can still be used as an example to estimate sample sizes.

In order to have some estimate of the variability between site-pairs, data from another pair of sites, K-1 and K-8, were used along with the data from P1 and SL-1. All of the data used appear in Appendix B. K-1

and K-8 are both 5th order stream sites on the Kawishiwi River and have similar water quality and physical characteristics. The covariance of drift density at these two sites is positive, thus they appear to be a good pair. Drift sampling was carried out more frequently at these four sites than at any other potentially similar pairs of sites, thus these appear to be the best sites to use for analyses of variance in drift data. Confidence intervals for mean densities and the logarithm of the ratio of the densities for K1 and K8 are shown in Figures 8 and 9.

An analysis of the drift data from these two pairs of sites (Appendix B) produces an estimate of S = 1.077 log units, which could enable the detection of a change of .40 log units, with 89 samples per treatment (t), (e.g. 3 years of preoperational data from 6 pairs of sites, 5 samples per year). The detectable change in the log of the ratio (control site density/ treatment site density), .40 log units, is equivalent to a 49% increase or 33% decrease in mean density at impacted (treated) sites. (A sample calculation translating changes in log units to % increase at the impacted site is given in Appendix B). Changes are calculated relative to the density expected at the "treated" site based on the density of the "control" site and the relationship between the two. Alternatively, a design of 4 site pairs, sampled 5 times per year for two preoperative years (t = 40) could detect changes of .60 log units, or 81% increase or 45% decrease in the mean total density at impacted (treated) sites. Obviously, sampling must be carried out for an equal time period during the operation of the facility. By using more complicated calculations, tradeoffs between various sampling designs can be considered (Snedecor and Cochran 1976, pp 528-534). However, biologically, it is desirable to sample for several years, both preoperationally and operationally.

If only a limited period of preoperational monitoring is carried out (e.g. "y" years) then the requirement of a balanced design would seem to limit testing for impacts to the first "y" years of operation. However, if monitoring continues throughout the operation, then data from any "y" years after operation begins can be compared to the preoperational data.

Determination of Number of Sample Replicates

The within-site variability does not enter the paired site analyses, because the paired site approach examines only the relationship between parameters at pairs of sites. Nonetheless, it is important that the investigators have some confidence in the means used in the analysis. One way to establish a "good" estimate of means is to specify the maximum desired width of the confidence interval for the mean. It should be remembered that the confidence intervals bracket an interval in which the mean value occurs with a given probability, if a large number of replicate samples are taken. Therefore increasing the number of replicates taken decreases the size of the confidence intervals for the mean, but would not reduce the range of values observed.

Means and 95% confidence intervals based on the t-distribution were calculated for all sites and all dates with more than two replicates for the following parameters: number of diatom taxa on glass slides, number of invertebrate taxa on Hester-Dendy samplers and in drift nets, Shannon-Wiener diversity index (log base 2) of diatoms and of invertebrates in both Hester-Dendy and drift samples, and total density of invertebrates in Hester-Dendy and drift samples. As before, a log transformation was used for the confidence intervals and means of total densities. Means

and 95% confidence intervals based on the t-distribution were also calculated for the relative abundance of all invertebrate functional groups and some taxonomic groups of invertebrates, and for total densities of diatoms and all periphyton cells. In the case of relative abundance data, the sampling could be considered binomial and the appropriate formulas for confidence intervals for proportions applied. However, examination of the data showed high variability between samples. A chi-square test indicated significant differences from sample to sample in more cases than would be predicted by chance. This reflects the fact that the conditions on samplers are never identical. Therefore, the normal distribution was used to calculate confidence intervals for these proportions, assuming that the differences between the slides are random (i.e. normally distributed, with mean 0.)

Examination of confidence intervals from all sites, dates, and parameters showed no obvious consistent patterns of variation, except that all parameters were highly variable, with variances at one site varying over time. Thus, without further analysis, it is impossible to generalize about the numbers of samplers needed to reduce the width of the confidence intervals. However, examination of data from a few sites is suggestive. Given the mean and sample variance for a parameter at a site on some date, one can use those values to estimate the number of samples needed to attain a 95% confidence interval of the form $\bar{x} + \underline{L}\ \bar{x}$, where L = a percentage, divided by 100, for that parameter at that site. The formula for the number of samples is:

$$n = \frac{4S^2}{(L\ \bar{x}\ )^2}$$ (Snedecor and Cochran 1967, p. 516)

where L times 100 is the percentage error we will accept. Thus the number of samples needed is proportional to the coefficient of variation squared,

$(S/\bar{x})$. For total density estimates, however, where the logarithm of the density is used to calculate confidence limits, it is more sensible to specify the actual width of the confidence interval (w) in log units. Thus an interval of the form $(\text{Ln } \bar{x} \pm 1/2)$ brackets the mean density within the interval $(.6\ \bar{x},\ 1.6\ \bar{x})$. Then the formula becomes

$$n = \frac{4S^2}{(W)},$$

and the number of samples is proportional to the sample variance, S^2, because the magnitude of W was set rather than specify the interval in terms of a percentage of the mean, $L\bar{x}$.

In order to compare the variability in different parameters and different data types, a few sites sampled frequently were selected. Calculations of numbers of replicates needed to provide 95% confidence intervals of the form $\bar{x} \pm 25\%\ \bar{x}$ for other parameters are shown in Tables 2-4. For intervals of the form $\bar{x} \pm 50\%\ \bar{x}$, the number of replicates would be reduced by a factor of 4; for density estimates, intervals of the form $\bar{x} \pm 1$ unit or $(0.37\ \bar{x},\ 2.7\ \bar{x})$ also require 1/4 as many replicates.

The number of replicates required was calculated for each site for all dates on which 3 or more replicates were analyzed. Tables 2-4 show, for each of the three types of sampling and a variety of parameters, the maximum number of replicates required for each site, as well as the mean number of replicates.

Table 2 suggests that for drift samples, there is less variability between replicates in the Shannon-Wiener diversity index than in the simpler index, number of taxa. Further analysis of the data is needed to confirm this. However, this does not necessarily mean that the Shannon-Wiener index is a better measure of changes in the system, but only that it is

less sensitive to the differences between replicates. The biological sig ·ˆicance of changes in this index is not known.

Tables 2-4 show that the variability between replicates is lower for diatoms sampled with glass slides than for invertebrates sampled with either Hester-Dendy samplers or drift nets. Data from fall, 1976 at K-1 were unusually variable, and calculations were done both with and without them. It appears that using 6 slides would provide reasonable confidence intervals for both the number of taxa, the Shannon-Wiener diversity index, and total diatom and cell densities. Such a sampling effort is quite feasible, whereas the numbers of drift nets and Hester-Dendy samplers needed would have to be very high (20 drift nets, 23 Hester-Dendy samplers) to approach the average numbers required at some sites. Thus on the basis of variability between replicate samples, diatoms seem better prospects for quantitative monitoring of stream communities than benthic invertebrates.

Comparisons Between Different Sampling Methods

The Regional Copper-Nickel Study used three different methods for quantitative stream sampling: drift nets and Hester-Dendy samplers for invertebrates and glass-slide samplers for diatoms. The literature on advantages and disadvantages of these methods is discussed in the Aquatic Biology Operations Manual (Regional Copper-Nickel Study 1977). Experience with the advantages and disadvantages of the 3 methods is summarized in Table 5.

Each method has distinct advantages and disadvantages. Overall, while drift samples are more costly than Hester-Dendy samples, drift nets seem to be preferable to Hester-Dendy samplers because they collect from a wide variety of habitats and are less subject to loss. It should be noted, though,

that to fully characterize the benthic invertebrate community, several methods should be employed.

Diatom sampling with glass slides is probably the easiest sampling method overall. The major disadvantages seemed to be loss of samplers and some bias toward species which have a propensity for glass slides (e.g. Achnanthes minutissima). As with benthic invertebrates, more than one method should be used to sample the periphyton community if a complete characterization is desired.

RECOMMENDATIONS

1. Use of Biological Monitoring

Biological monitoring can be used in two ways. First it serves as insurance against non-detection of gross, catastrophic changes in biological systems. Second, it can detect smaller but statistically significant changes in biological systems, by intensive sampling over fairly long periods of time. (The paired-site approach can fulfill either purpose, depending on the effort expended.)

Physical and chemical monitoring must precede biological monitoring so that similar sites can be paired. Key parameters to be matched for streams are flow, pH, temperature, alkalinity, and streamside vegetation. This should be supplemented by enough paired-site biological monitoring so that at a minimum, catastrophic changes will be detected. Catastrophic changes might, for instance, be defined as changes in ratios of drift densities of 1.5 log units. This is equivalent to a decrease of 78% or increase of 348% in density at the treated site relative to expected values at that site. Based on the illustrative example from Study drift data, such a program would require a sample size of 6, which might be achieved by 3 pairs of sites sampled once a year for two years preoperationally. This estimate of sample size should not use the formulas discussed in Appendix B, as the normal approximation is not accurate for such small samples. Exact power calculations must be made.

Biological monitoring to detect smaller changes should be undertaken in two kinds of situations.

1) Situations where some species are of particular importance in themselves

or in relation to important species; or

2) Situations where data on long term impacts of new types of development are lacking (e.g. when new chemicals are released).

For example, if acidified streams are a potential problem, there is already a body of existing information on a wide range of organisms which shows that pH levels of less than or equal to 5.0 have short-term biological effects and levels of 6.0 or less have produced long-term effects. On this basis, it is more efficient to monitor pH continuously than to detect significant changes in biological parameters, which will take years. If there were no data available on the effects of pH, then it would be important to institute a long-term monitoring program around a source of acid.

2. Taxonomy

Even with the paired-site approach, it is clear that substantial biological monitoring efforts are required to detect non-catastrophic change. Therefore, sampling and taxonomic analysis should not be carried out at any more detailed level than can reasonably be considered in data analysis. It appears necessary to identify individual species only when a particular species is considered important or when such identification is necessary in order to define larger groups, such as invertebrate functional groups or Husted's (1937-1938) diatom groups. As Eberhardt (1976) points out, in most cases we do not know which species are critical to the structure and function of ecological systems, thus we can rarely select such critical species for monitoring. McKenzie et al. (1977) discuss in detail what appears to be the present state of the art of impact analysis in terms of the use of taxonomic and physical/chemical data.

3. Number of samples

Since preliminary estimates of sample sizes required to achieve a test of a desired power may be in error, it may be desirable to sample more frequently than indicated by preliminary estimates. The extra samples can be saved. After several years of sampling, if variances are higher than expected and the analysis of variance is less powerful than required, additional samples could be analyzed and added to the analysis.

4. Methods recommended in the Study Area

The first consideration in choosing methods of monitoring is the choice of "important" functions or components to monitor. Once these are chosen on the basis of economic or aesthetic considerations, known sensitivities to stress, or ecological importance, then methods of sampling can be considered. If studies of stream invertebrates are selected, then drift nets appear preferable to Hester-Dendy samplers because they are rarely lost and are less vulnerable to problems due to fluctuating flows and water levels. If no particular component is known to be more important than others, and the objective is to choose one of the many important biological components of the system, then periphyton sampling appears to provide the most reliable estimates of parameters for a reasonable sampling effort. To achieve similar reliability from drift net and Hester-Dendy samplers requires numbers of samples that are impractical. However, it is not known to what extent changes in periphyton reflect the many possible kinds of change in ecosystems. Therefore, if other components are of particular importance, the substitution of periphyton studies for other kinds of studies is not recommended.

SIGNIFICANCE OF CHANGE

One of the major problems in designing biological monitoring programs is ignorance of what constitutes significant change in the functioning of ecosystems. It is desirable to monitor the "health" of the system, but there is no consensus among ecologists on how this can be measured. The relation of diversity to the health of an ecosystem has been discussed for years, but does not appear appropriate for all kinds of systems. At a recent workshop on "The Biological Significance of Environmental Impacts" Buffington (1975) summarized as follows:

> "An impact is significant if it results in a change that is measurable in a statistically sound sampling program and if it persists, or is expected to persist, more than several years at the population, community, or ecosystem level."

The point is that even with a well-designed sampling program, the impacts that are detected are not subtle, and are thus likely to be significant in terms of the functioning of the system. All of the data from the Regional Copper-Nickel Study documenting natural variability seem to support this position; if anything is statistically significant in these highly variable ecosystems,it is probably biologically significant. The impossibility of adequately monitoring every facet of the ecosystem requires careful consideration of the design and planned utilization of every monitoring program before it is initiated.

Literature Cited

Buffington, J.D. 1975. A synthetic definition of biological significance. Pages 319-327, in The Biological Significance of Environmental Impacts. Proc. of NRC Workshop, Buffington, T.T. McFadden and R.K. Sharma, eds. (Available throught NTIS)

Cairns, J.Jr. 1976. Summary. Pages 235-242. in J. Cairns Jr., K.L. Dickson, and G.F. Westlake, eds. Biological Monitoring of Water and Effluent Quality. Proc. of Symp. A.S.T.M. Philadelphia.

Cairns, J. Jr. 1973. Rapid biological monitoring systems for determining aquatic community structure in receiving systems. Pages 148-163. in J. Cairns, Jr. and K.L. Dickson, eds. Biological Methods for the Assessment of Water Quality. Proc. of Symp. A.S.T.M. Philadelphia.

Cochran, W.G. and G.C.Cox, 1957. Experimental Designs. John Wiley and Sons, New York. 611 pp.

Cooke, R.Dennis, and K. Larntz, 1973. Sample size determination for "Fixed Effects" ANOVA Models. Technical Report # 212. Dept. Applied Statistics, U. Minnesota, St. Paul, Mn.

Eberhardt, L.L. 1976. Quantitative ecology and impact assessment. J. Env. Manage. 4:27-70.

_____. 1978. Appraising variability in population studies. J. Wildl. Manage. 42: 207-238.

Gaufin, A.R. 1973. Use of aquatic invertebrates in the assessment of water quality. Pages 96-116, in J. Cairns, Jr. and K.L. Dickson, eds. Biological Methods for the Assessment of Water Quality. Proc. of Symp. A.S.T.M. Philadelphia.

Gorham, E. and A.G. Gordon, 1960. The influence of smelter fumes upon the chemical composition of lake waters near Sudbury, Ontario, and upon the surrounding vegetation. Can. J. Botany 38: 477-487.

Herricks, E.E. and V.O. Shanholtz. 1976. Predicting the environmental impact of mine drainage on stream biology. 1976. Trans. ASAE. 271-274.

Hirsh, A. 1976. The baseline study as a tool in environmental assessment. Presented at AIBS Symposium on Biological Evaluation of Environmental Impact, New Orleans, June 1976. In press CEQ.

Hustedt, F. 1937-1938. Systematische und Ökologische Untersuchungen uber die Diutomeenflora von Java, Bali, und Sumatra. (Systematic and ecological investigation of the diatom flora of Java, Bali, and Sumatra.) Arch. f. Hydrobiol., Supplement-Band. (Stuttgart) 15: 131-177; 16:187-295; and 16:393-506.

Kaesler, L., J.C. Cairns, and J.S. Crossman. 1974. Redundancy in data from stream surveys. Water Research. 8:637-642

Likens, G.E., F.H. Bormann, and N.M. Johnson. 1969. Nitrofication: Importance to nutrient losses from a cutover forested ecosystem. Science. 163:1205-1206.

McKenzie, D.H., L.D. Kannberg, K.L. Gore, E.M. Arnold, and D.G. Watson. 1977. Design and analysis of aquatic monitoring programs at nuclear power plants. PNL-2433/NRC 10. Prepared for the Nuclear Regulatory Commission. Battelle Pacific Northwest Laboratories, Richland, Washington. (available through NTIS) 126 pp.

Morgan, W.S.G., 1976. An electronic system to monitor the effects of changes in water quality on fish opercular rhythms. Pages 38-55 in J. Cairns Jr., K.L. Dickson and G.F. Westlake, eds. Biological monitoring of water and effluent quality, Proc. of Symp. ASTM Philadelphia.

Mount, D.I. 1976. Biotic monitoring, pages 20-23, in J. Cairns, Jr., K.L. Dickson and G.F. Westlake, eds. Biological monitoring of water and effluent quality, Proc. of Symp. ASTM Philadelphia.

Poels, C.L.M. 1976. An automatic system for rapid detection of acute high concentrations of toxic substances in surface water using trout. pp. 85-95 in J. Cairns Jr., K.L. Dickson, and G.F. Westlake, eds. Biological monitoring of water and effluent quality, Proc. of Symp. ASTM Philadelphia.

Regional Copper-Nickel study. 1978a. Benthic Invertebrates of lakes. Minnesota Environmental Quality Board, St. Paul, MN.

_____.1978b. Lake zooplankton. _____.

_____. 1978c. Lake phytoplankton. _____.

_____. 1978d. Lake fisheries report. _____.

_____. 1978e. Stream benthic invertebrates. _____.

_____. 1978f. Stream periphyton. _____.

_____. 1978g. Stream fish. _____.

_____. 1978h. The toxicity of heavy metals, beneficiation reagents. and hydrogen ion to aquatic organisms. _____.

_____. 1978i. Erie Mining Project - Biological sampling. _____.

_____. 1977. Operations Manual, Aquatic Biology. _____.

Snedecor, G.W. and W.G. Cochran. 1967. Statistical methods. Iowa State University Press. Ames, Iowa. 573 pp.

Sokal, R.R. and Rohlf, F.J. 1969. Biometry. W.H. Freeman and Co., San Francisco.

Stokes, P.M., T.C. Hutchinson, and K. Krauter. 1973. Heavy metal tolerance in algae isolated from polluted lakes near the Sudbury, Ontario smelters. Water Pollution Research in Canada 8:178-201

Van Winkle, W. 1977. Conclusions and recommendations for assessing the population-level effects of power plant exploitation: The optimist, the pessimist, and the realist. pages 365-372. in Van Winkle, ed. Proceedings of the conference on assessing the effects of power plant induced mortality on fish populations. Pergamon Press, N.Y.

Table 1. Summary of some results from McKenzie et al (1977), analysis of data from aquatic biological monitoring at nuclear power plants.

SITE AND ORGANISMS	LAKE MICHIGAN BENTHOS ZION PLANT			HADDAM NECK BENTHOS		LAKE MICHIGAN PHYTOPLANKTON ZION PLANT
no. of station pairs	10			2		6
sampling frequency per year	6			9		12
no. of years	4			5		5
preoperational	2			3		2
operational	2			2		3
no. of paired samples	240			90		350
model and data transformaton	mult*-log	additive $\sqrt{}$	additive none	mult* log	additive $\sqrt{}$	mult-* log
α level	10%	10%	10%	10%	10%	10%
power of test (1-β)	80%	80%	80%	80%	80%	80%
least detectable diff.	.25 log units	10.7 5 units		.356 log		0.20
% increase	28%		30%	42%		20%
% decrease	22%		30%	29%		18%
difference found?	yes	no	no	yes	yes	no

*mult. refers to multiplicative models which lead to the use of a log transform.

Table 2. Mean and maximum numbers of sample replicates estimated (for all dates sampled) as needed to provide 95% confidence limits of specified width for mean values of parameters of replicate drift samples at a few sites sampled by Regional Copper-Nickel Study. The number of replicates was calculated for intervals of half width equal to 25% of the mean for all parameters except total densities. For these, calculations were based on log transformed data and intervals were set at half width equal to .5 log units.

DRIFT REPLICATES

	MEAN NUMBER			MAXIMUM NUMBER		
SITE PARAMETER	P-1	K-1	P-5	P-1	K-1	P-5
No. of taxa	9.0	10.3	3.0	35.2	35.2	5.2
Shannon-Wiener diversity	1.6	4.4	1.8	12.9	18.3	6.6
Relative abundance of collectors-gatherers	2.7	16.9	8.9	7.1	44.0	15.7
Total density invertebrate	8.9	19.9	5.2	25.4	58.1	13.6

Table 3. Mean and maximum numbers of sample replicates estimated (for all dates sampled) as needed to provide 95% confidence limits of specified width for mean values of parameters of replicate Hester-Dendy samples at a few sites sampled by Regional Copper-Nickel Study. The number of replicates was calculated for intervals of half width equal to 25% of the mean for all parameters except total densities. For these, calculations were based on log transformed data and intervals were set at half width equal to .5 log units.

	HESTER - DENDY REPLICATES					
	MEAN NUMBER			MAXIMUM NUMBER		
SITE / PARAMETER	P-1	P-5	SL-1	P-1	P-5	SL-1
No. of taxa	4.9	2.2	5.6	12.7	7.54	20.1
Shannon-Wiener diversity	1.9	12.9	5.5	3.3	32.5	20.2
Relative abundance of collectors-gatherers	16.5	23.8	20.1	77.8	78.9	60.4
Total density of invertebrates	4.0	7.4	17.5	13.9	56.1	88.5

Table 5. Mean and maximum numbers of sample replicates estimated (for all dates sampled) as needed to provide 95% confidence limits of specified width for mean values of parameters of replicate periphyton galss slide samples at a few sites sampled by Regional Copper-Nickel Study. The number of replicates was calculated for intervals of half width equal to 25% of the mean for all parameters except total densities. For these, calculations were based on log transformed data and intervals were set at half width equal to .5 log units. Values in parentheses (*) were calculated omitting data from one date at K-1.

PERIPHYTON SLIDE REPLICATES

	MEAN NUMBER				MAXIMUM NUMBER			
SITE PARAMETER	P-1	K-1	P-5	SL-1	P-1	K-1	P-5	SL-1
No. of taxa, diatoms	3.5	2.1	6.4	5.7	10.8	5.5	16.0	14.1
Shannon-Wiener diversity,diatoms	4.6	1.6	3.0	4.0	13.7	3.8	7.2	17.1
Total density of diatoms	2.1	60.0 (3.4)*	4.3	2.0	7.0	230 (9.1)*	9.3	5.7
Total density of all cells	2.2	91.2 (3.3)*	4.1	1.7	7.1	355 (8.7)*	9.0	5.7

Table 5. Comparison of quantitative aquatic biological sampling methods employed during the Regional Copper-Nickel Study.

	DRIFT	HESTER/DENDY	PERIPHYTON(taxonomic analyses)*
Time			
Sampling ($\leq$ 6 reps)	1 hr/site	1/4 hr/site	1/4 hr/site
processing	4-8 hrs/sample	2-4 hrs/sample	1/6 hr/sample**
analysis	6-7 hrs/sample	2-2½ hrs/sample	1½-2 hr/sample
Ease of Sampling	—	+	+
Sampling problems	1. moving water necessary to sample 2. affected by factors such as life cycles, night length, moon, weather, flow 3. generally large samples with large amounts of detritus causing difficult processing 4. specimens often become battered in the net 5. current velocity measurements are necessary but often difficult	1. samplers frequently lost because of vandalism or changes in flow condition 2. samples generally small 3. water level fluctuations during colonization period may subject samplers to varying conditions and in somes cases expose them 4. specimens often damaged when samplers are scraped 5. limited number of taxa collected	1. samplers frequently lost because of vandalism or change in flow conditions 2. water fluctuation during exposure often subjects samplers to varying conditons 3. it is more difficult to develop a reference collection or have specimens verified than with insects. 4. involved processing procedures
Sampling advantages	1. very few samples lost 2. large sample size in general 3. adaptable to changing conditions 4. large number of taxa collected	1. quick and easy field sampling 2. samples easy to process	1. quick and easy field sampling 2. easy storage of samples

Table 5. continued

	DRIFT	HESTER/DENDY	PERIPHYTON (taxonomic analyses)*
Sampling bias	1. emerging insects	1. scrapers and collectors over estimated 2. few groups collected	1. Achnanthes and Cocconeis prefer glass slides and can crowd other species

* species proportional counts

** assumes at least 10 replicates processed simultaneously

Figure 1a. Mean diatom diversity (Shannon-Wiener Index, $\log_2$) on three glass slides P-1, and annual means based on four dates per year.

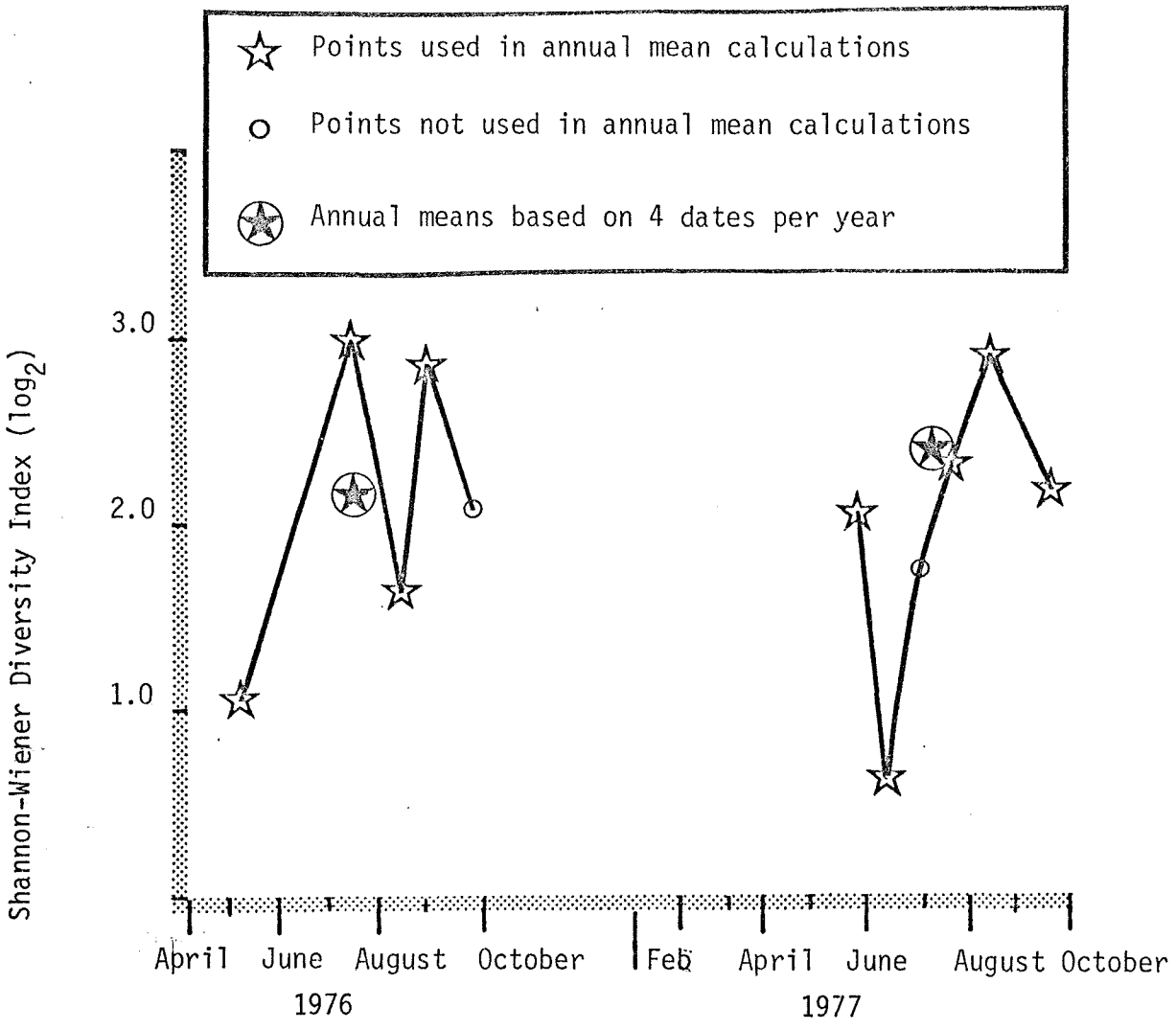


Figure 1b. Mean and 95% confidence intervals for mean diversity at P-1.

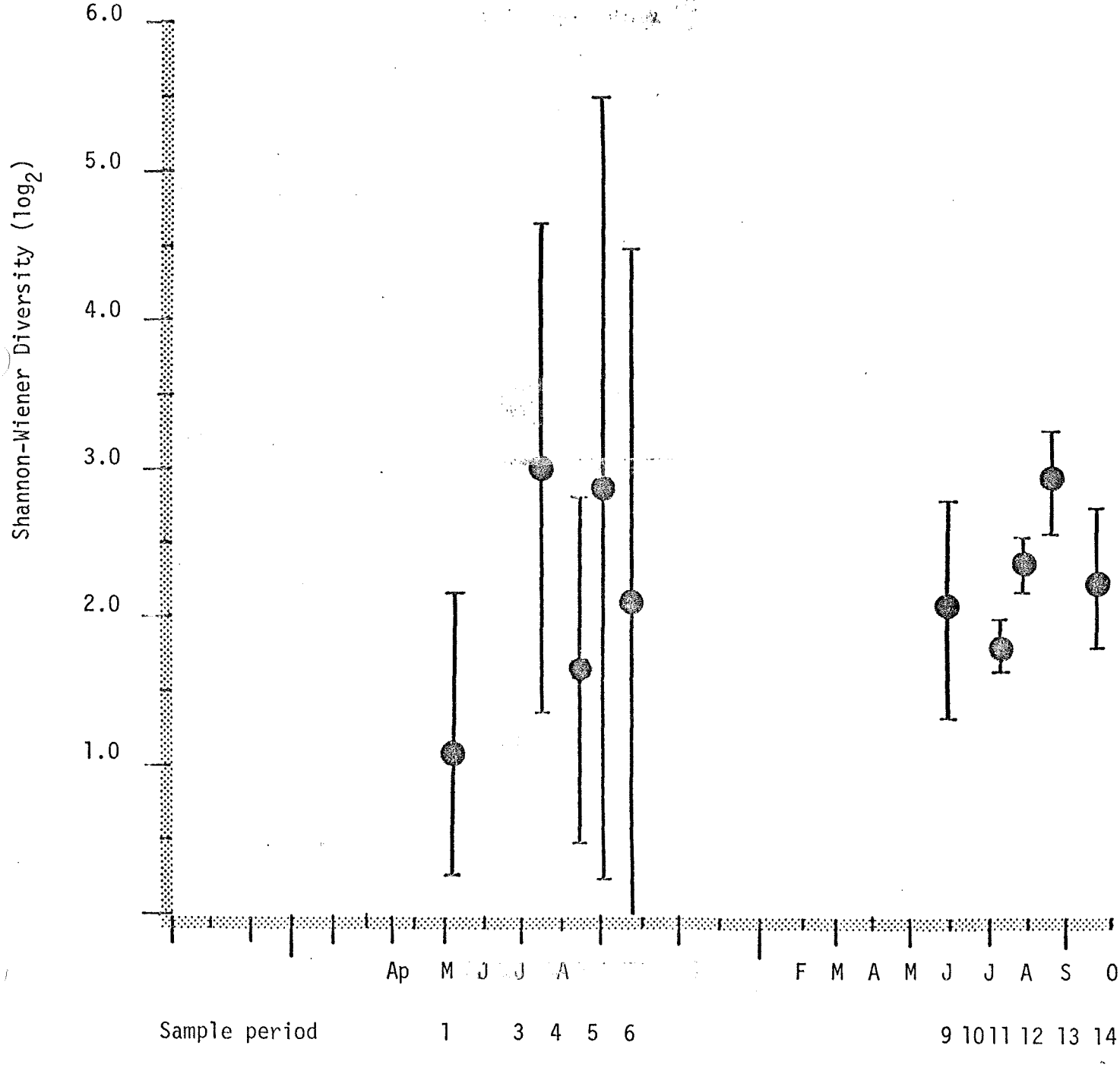


Figure 1c. Temperature and discharge at P-1 between October, 1975, and October, 1977.

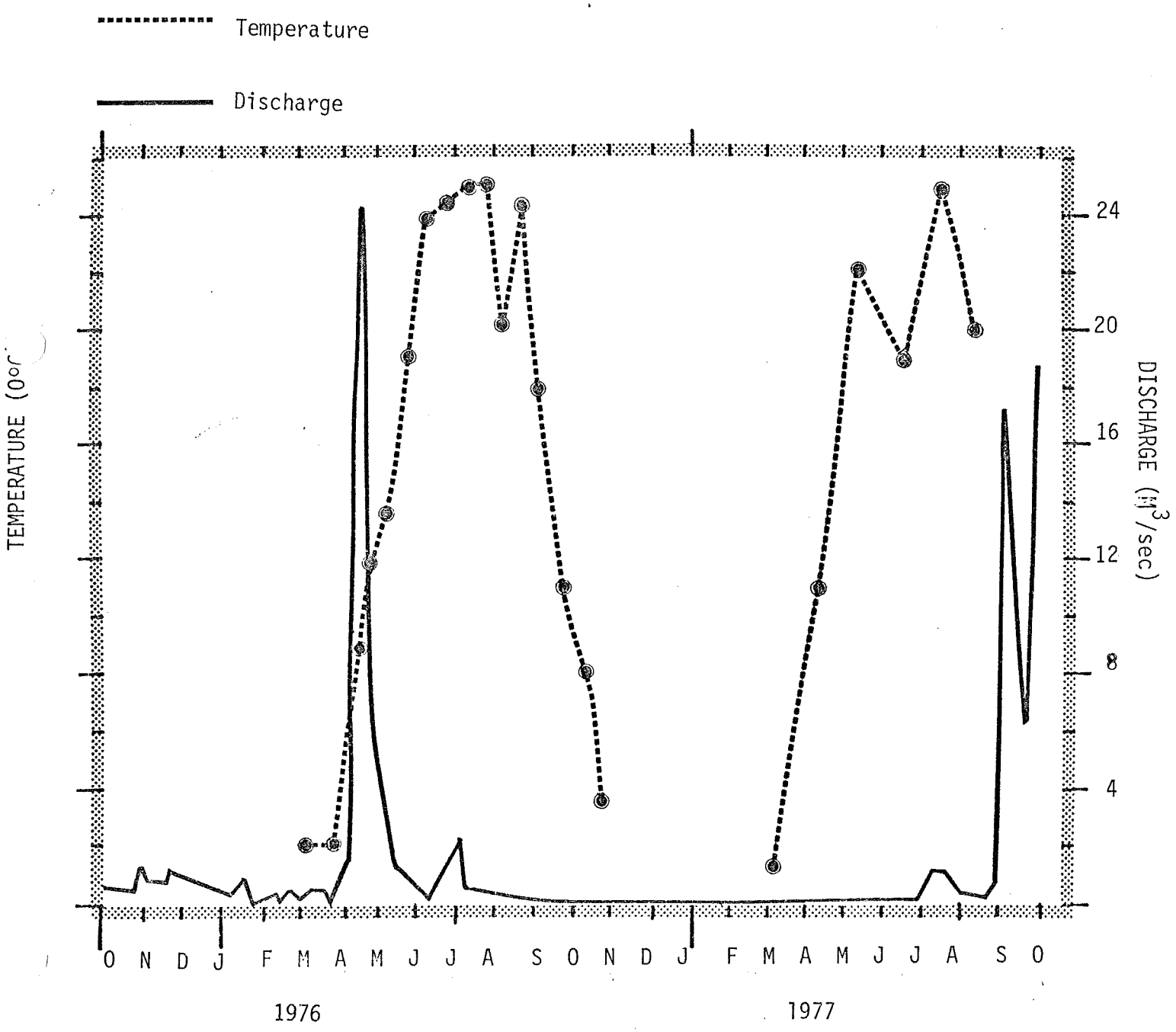


Figure 1d. Annual mean diatom diversity at station P-1

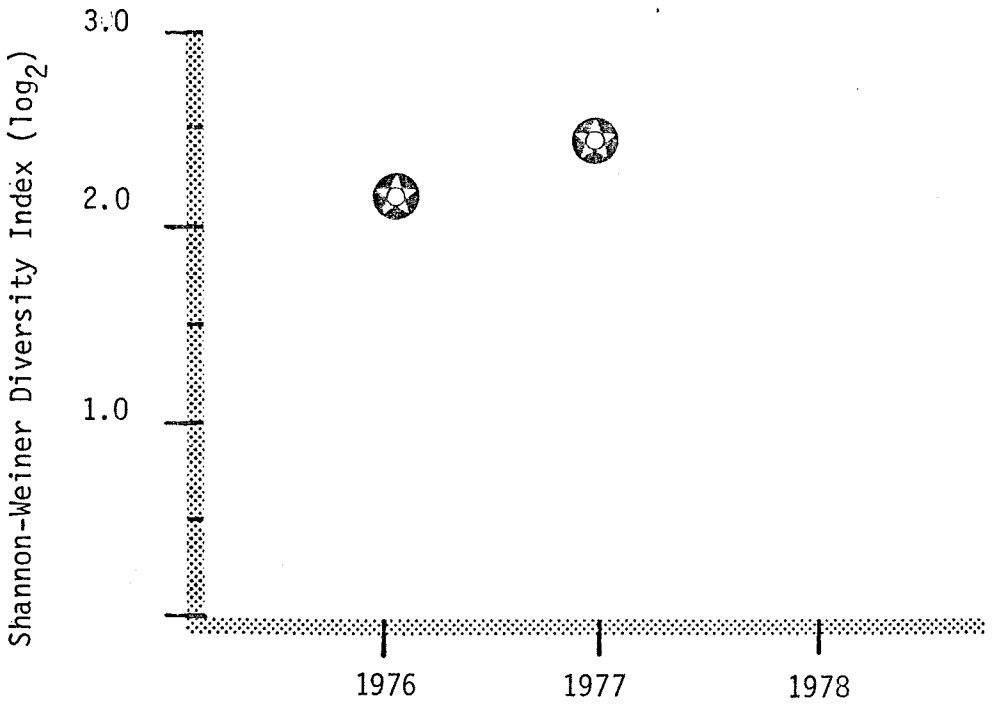


Figure 2. Frequency distribution of mean Shannon-Wiener Diversity Index for Diatoms on glass slides, for all sites and dates sampled by Regional Copper-Nickel Study.

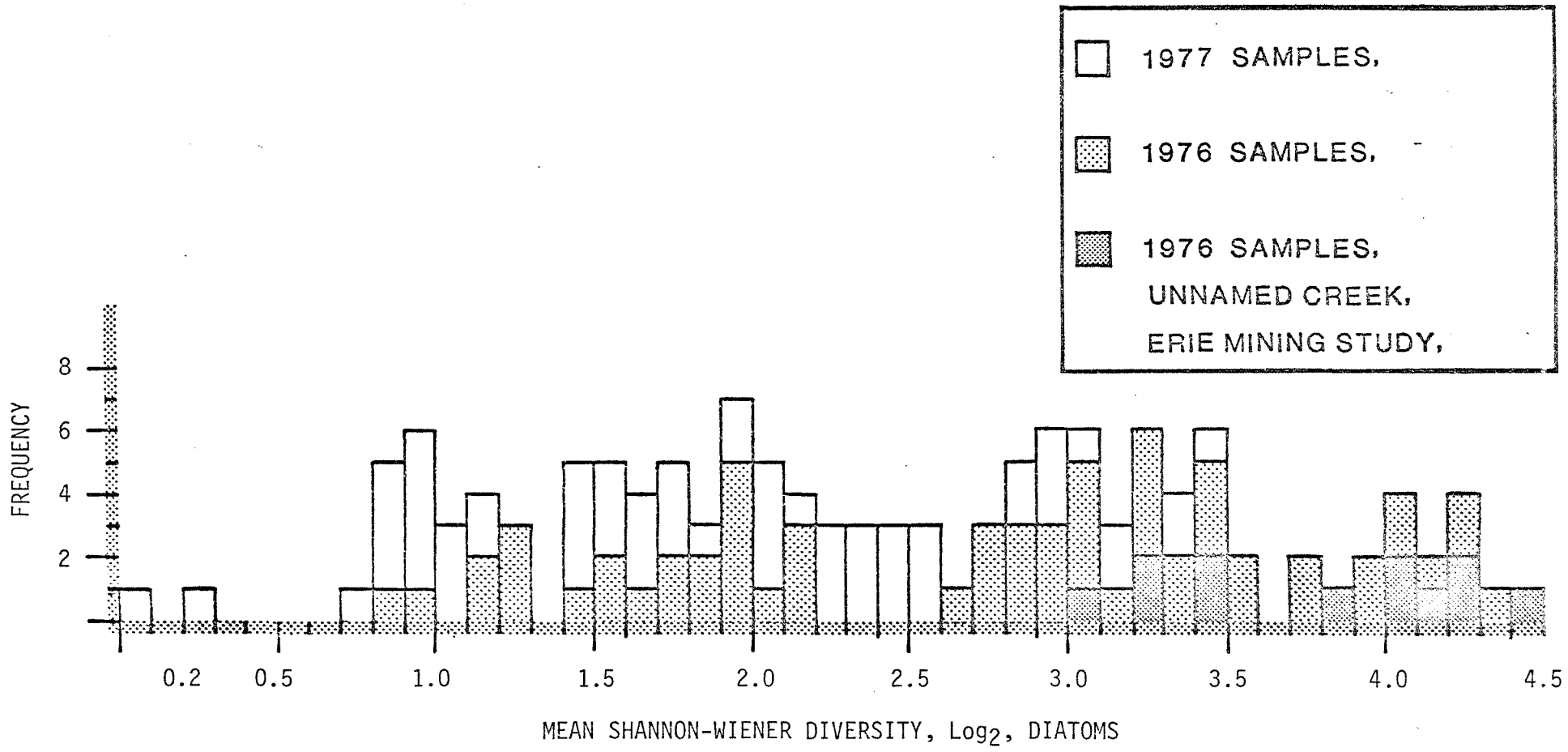


Figure 3. Frequency distribution of mean number of taxa of diatoms on glass sites, for all sites and dates sampled by Regional Copper-Nickel Study.

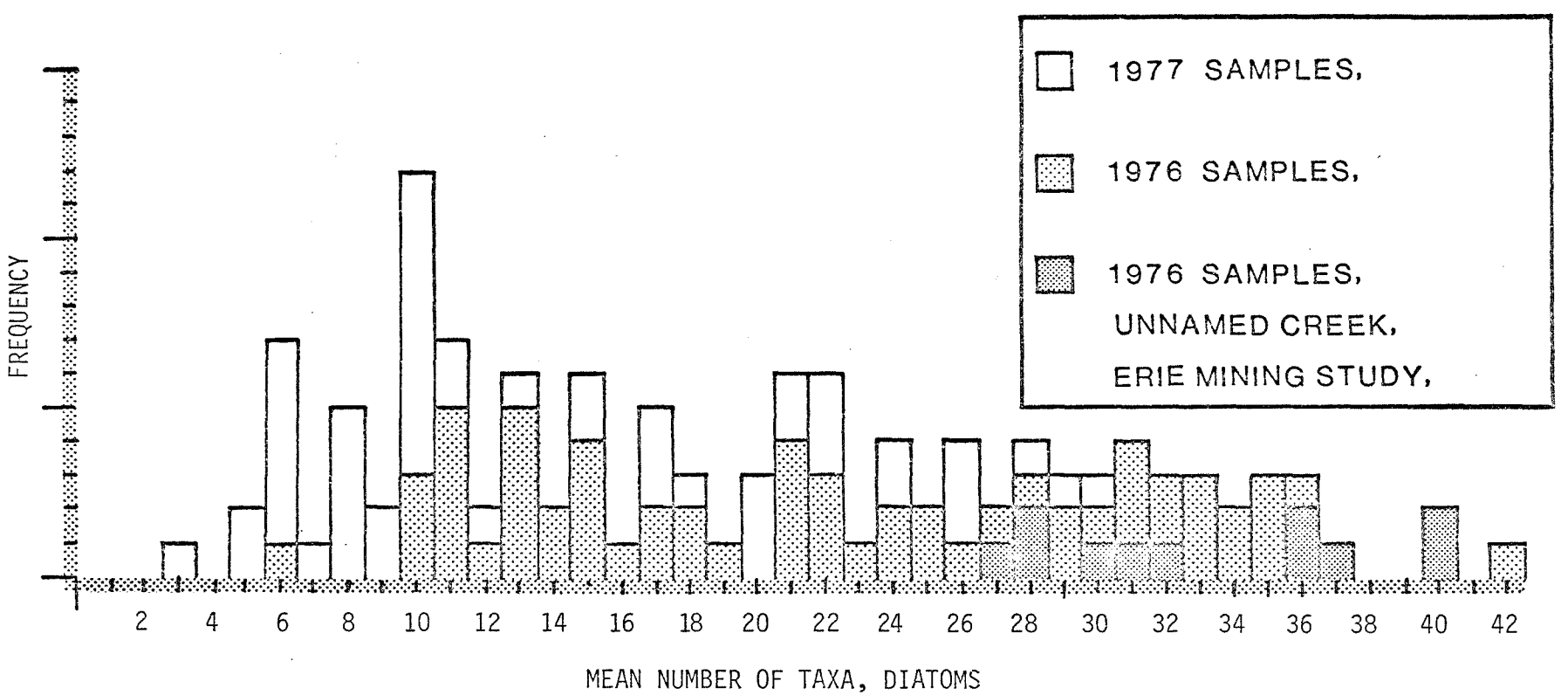


Figure 4. Frequency distribution of mean number of taxa in drift samples, for all sites and dates sampled by Regional Copper-Nickel Study.

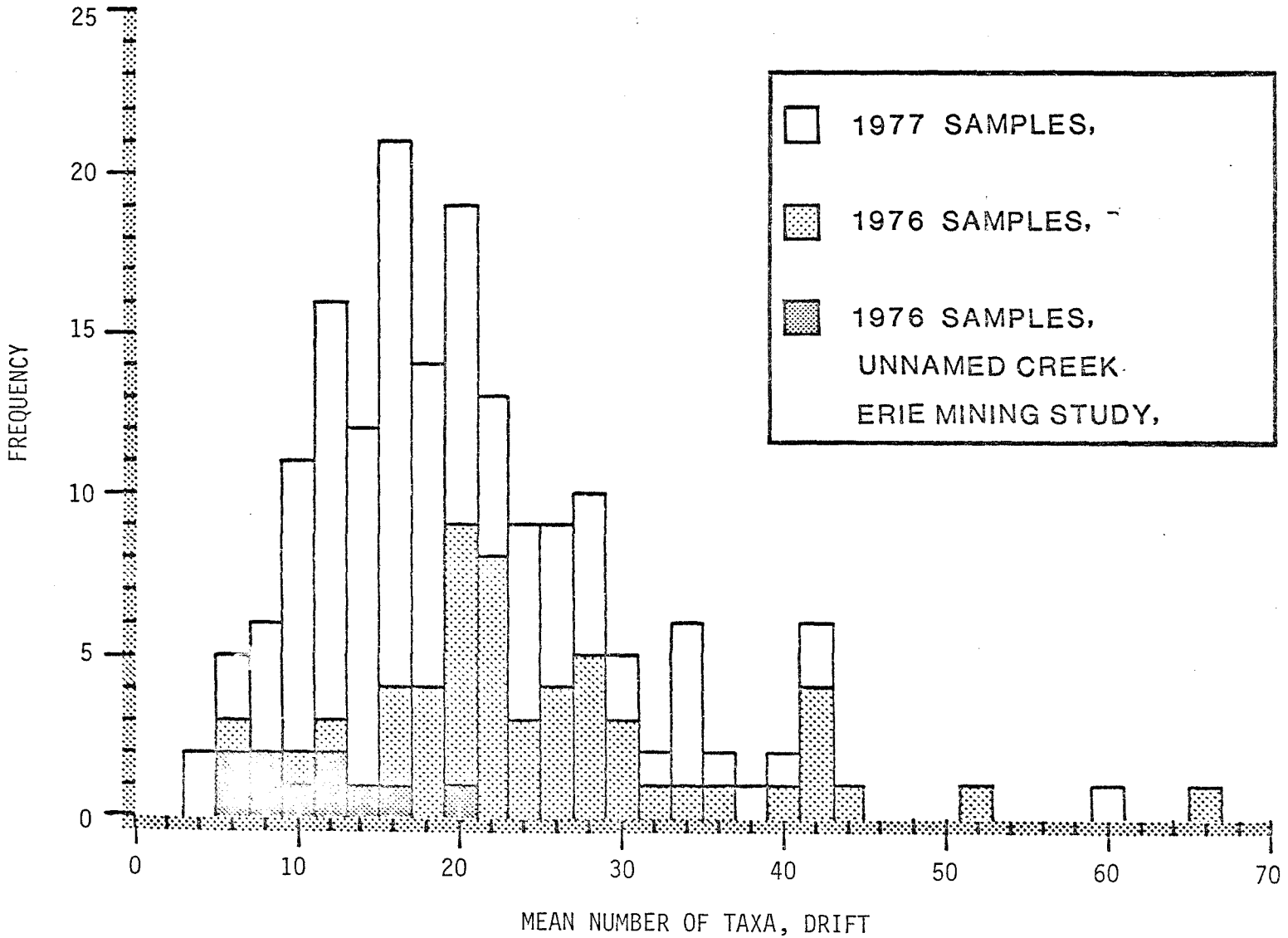


Figure 5. Frequency distribution of geometric means of total numbers of invertebrates in drift samples, for all sites and dates sampled by Regional Copper-Nickel Study.

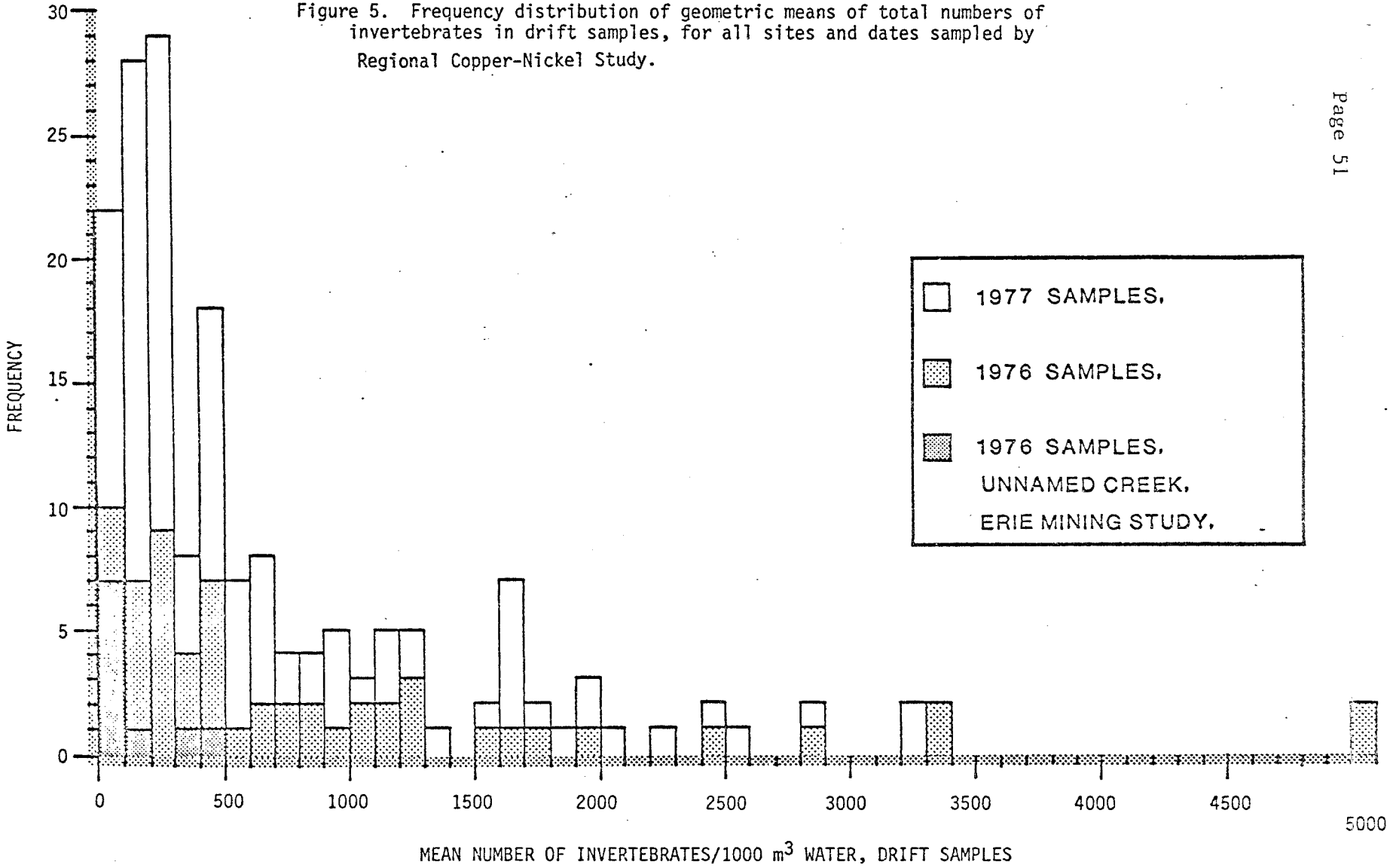


Figure 6. Means and 95% confidence intervals (based on log transform) for number of organisms/1000 m^3 water in drift samples at stations SL-1 and P-1. Means are based on three drift nets.

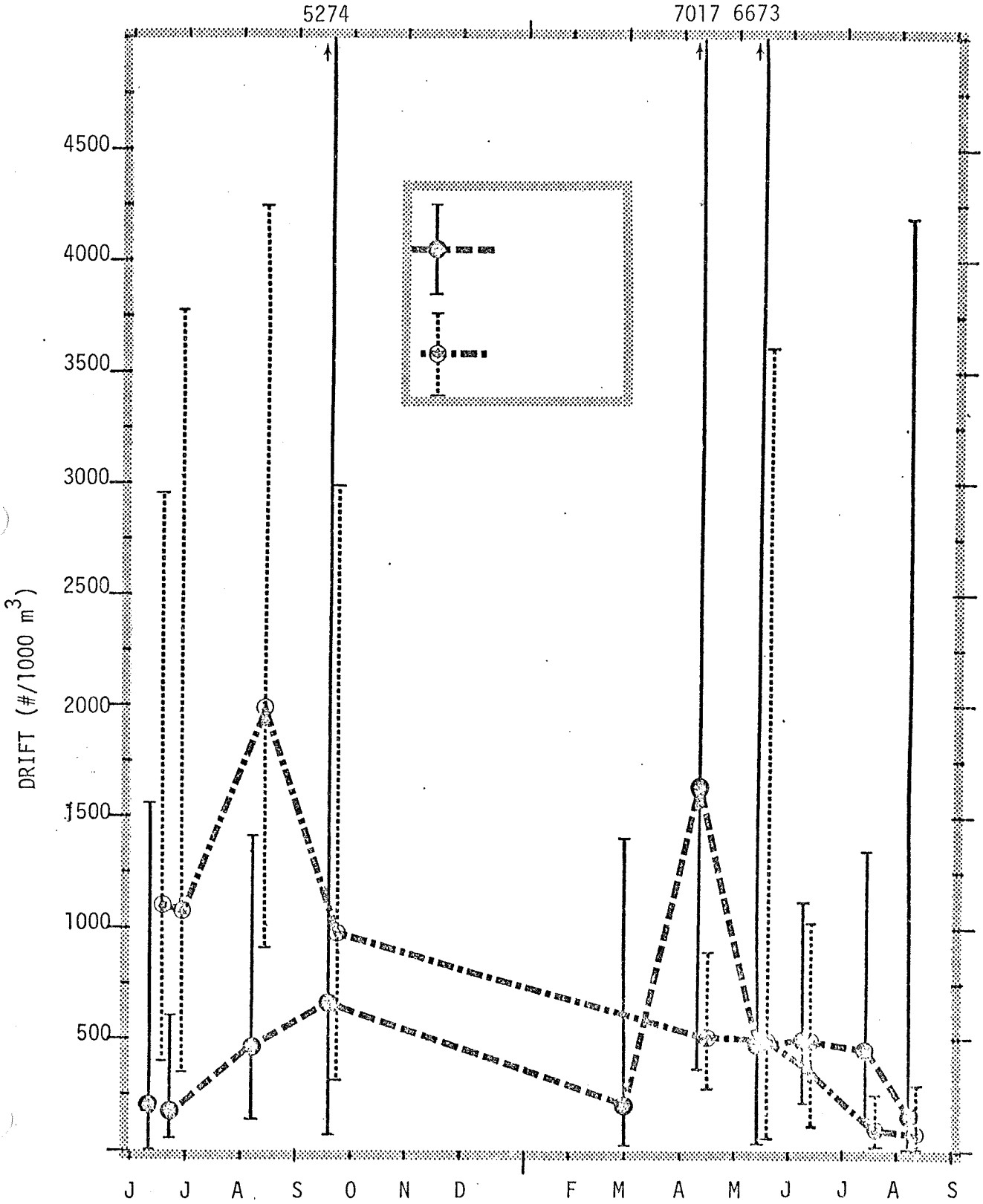


Figure 7. Natural log of ratio of drift densities at sites P-1 and SL-1 (ln P-1/SL-1) in 1976 and 1977.

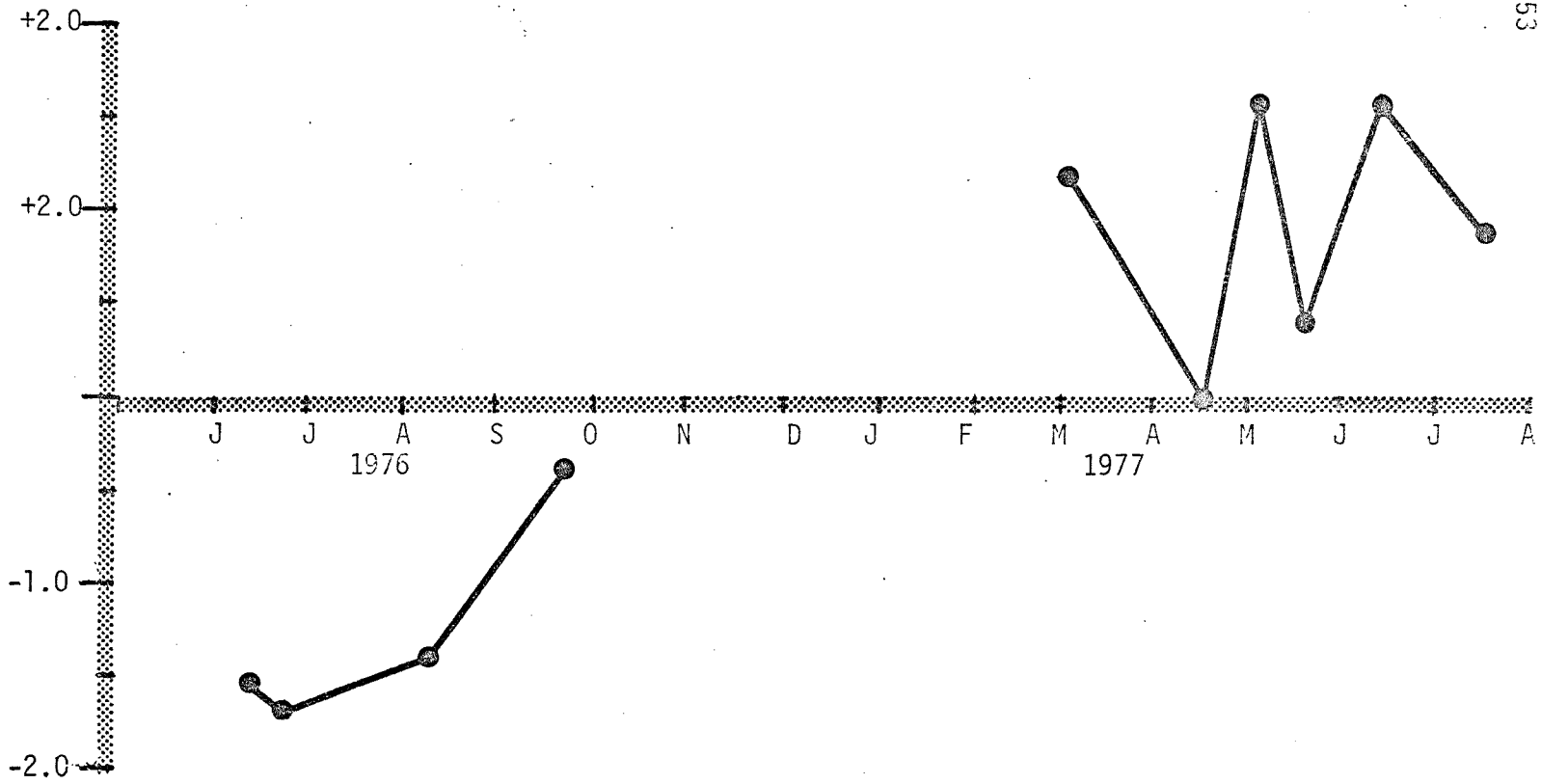


Figure 8. Means and 95% confidence internals (based on log transform) for number of organisms/1000 m^3 water in drift samples for stations K-1 and K-8. Means are based on 3 drift nets.

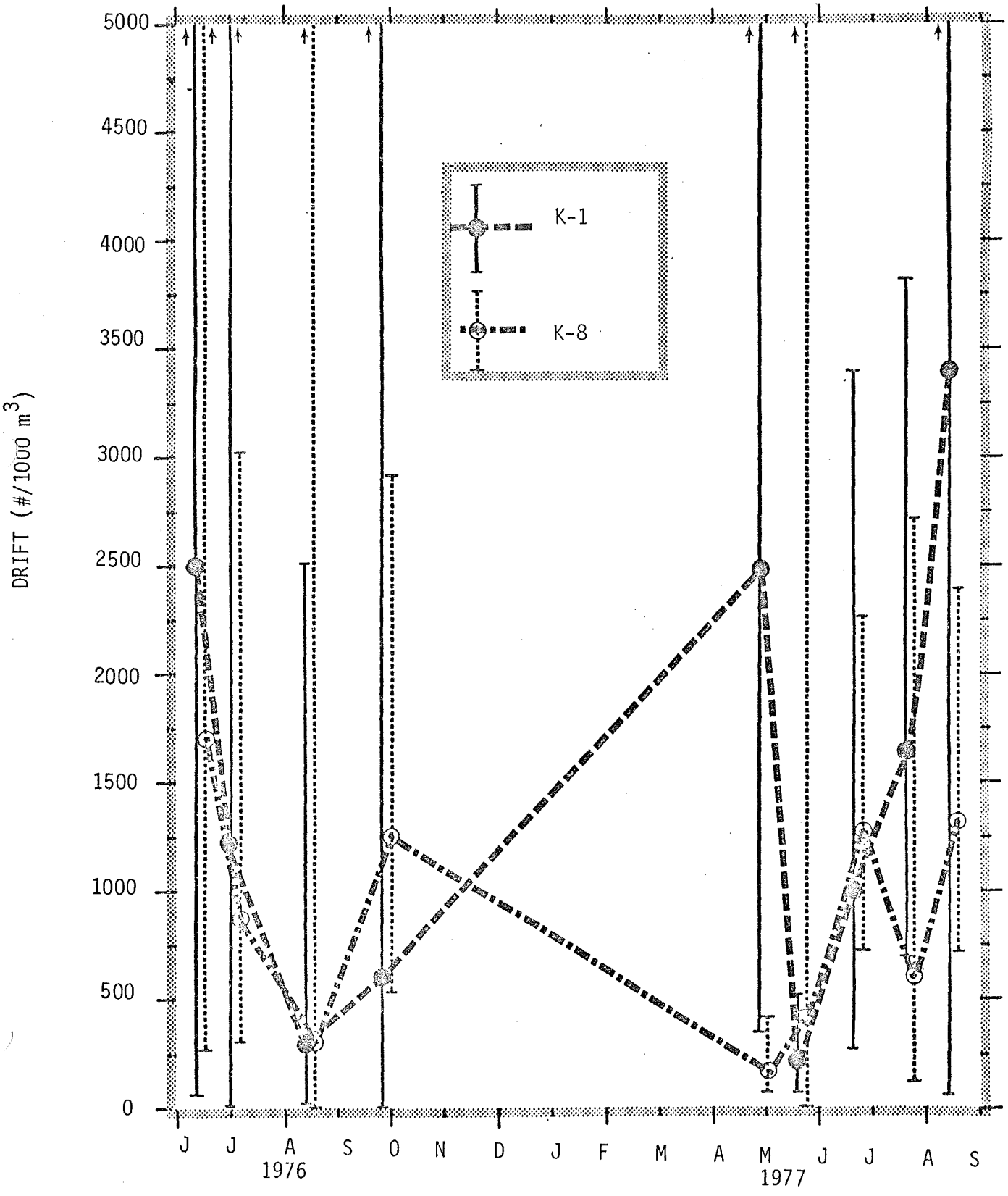
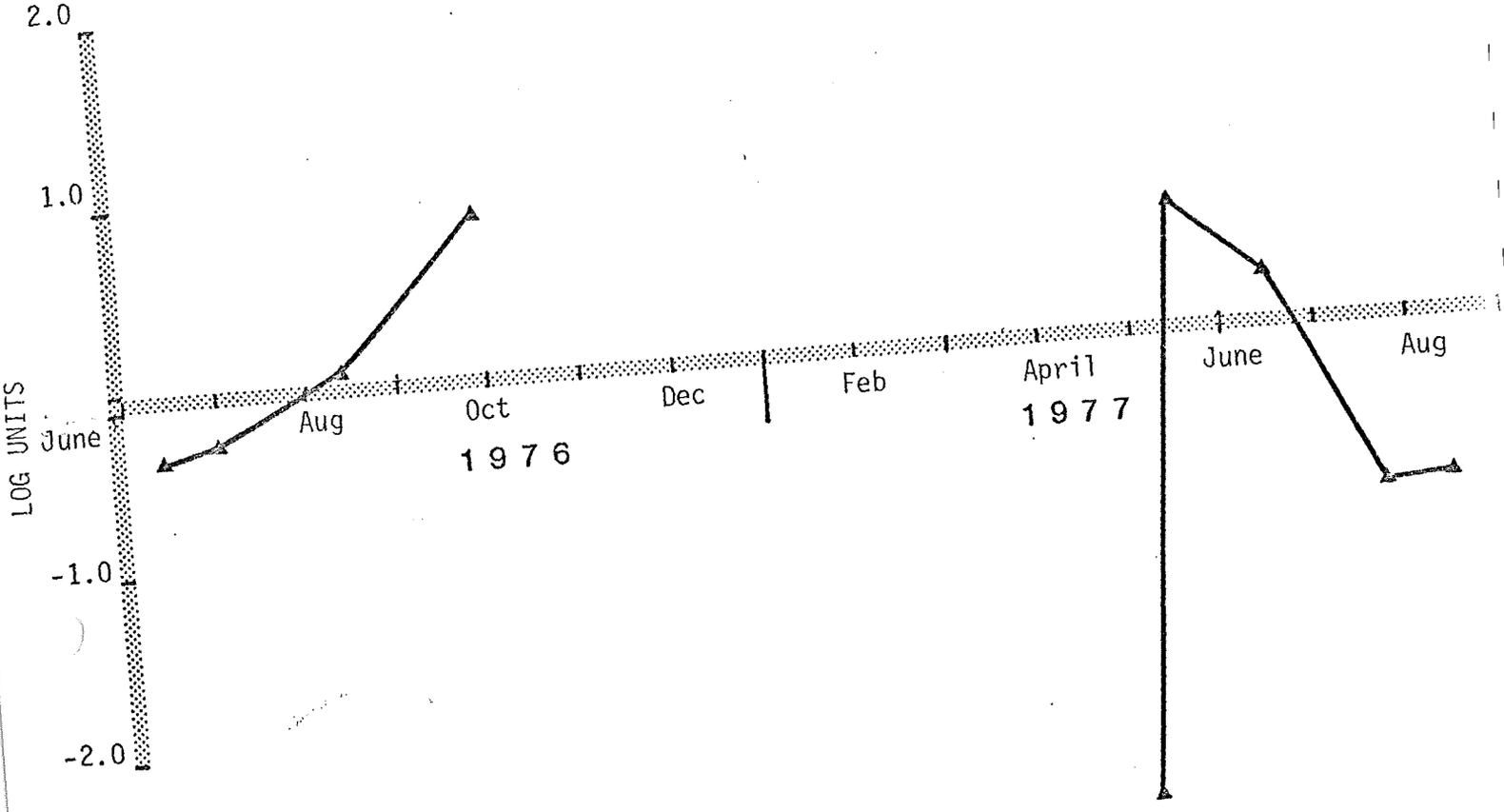


Figure 9. Natural log of the ratio of drift densities at stations K-1 and K-8 (ln K-8/K-i) in 1976 and 1977.



APPENDIX A.

The following tables show the number of samples of each type collected at each station for all sample periods. The dates included in each sampling period are shown below.

Sample Period	Dates
1	May 1, 1976 - June 18, 1976
2	June 28, 1976 - July 15, 1976
3	July 19, 1976 - July 30, 1976
4	August 9, 1976 - August 20, 1976
5	August 30, 1976 - September 10, 1976
6	September 11, 1976 - October 15, 1976
7	February 1, 1977 - March 31, 1977
8	April 1, 1977 - May 15, 1977
9	May 16, 1977 - May 31, 1977
10	June 1, 1977 - June 25, 1977
11	June 27, 1977 - July 15, 1977
12	July 18, 1977 - July 31, 1977
13	August 1, 1977 - August 26, 1977
14	August 27, 1977 - November 5, 1977

Site names followed by the letters LA designate samples taken later but in the same sample period indicated.

A-1. Number of Samples, Diatom Slides, 1976

SITE	DATE				
	1	3	4	5	6
F1	3	0	3	0	3
BR1	3	0	3	0	3
KC1	3	0	3	0	3
P5	3	0	3	0	3
K2	3	0	3	0	3
BI1	1	0	3	0	3
D1	3	3	3	0	1
E1	3	3	3	3	3
SL3	3	0	3	0	3
K8	3	3	0	3	3
K1	3	3	3	3	3
R1	3	3	0	3	0
SR3	3	0	1	0	3
K5	3	0	3	0	0
P1	3	3	3	3	3
P2	3	0	3	0	3
SL1	3	3	3	3	3
EM1A	0	0	6	6	6
EM1	0	0	6	6	6
EM3	0	0	6	6	6
EM1LA	0	0	0	0	5
EM3LA	0	0	0	0	6
EM1ALA	0	0	0	0	6

A-2. Number of Samples, Diatom Slides, 1977

SITE	DATE					
	9	10	11	12	13	14
SP1	0	2	0	2	0	0
LI3	0	2	0	2	0	0
LI1	0	2	0	2	0	0
LI2	0	2	0	0	0	0
P5	2	2	2	2	2	0
K2	0	0	0	2	2	0
BI1	0	2	0	2	0	0
D1	4	0	0	4	4	2
SC1	0	2	0	2	0	0
SE1	0	2	0	0	0	0
E1	3	0	0	3	3	0
SL2	2	0	0	0	1	0
K8	3	0	0	4	4	0
K1	4	0	0	4	0	0
SR2	4	0	0	4	0	2
SR3	2	0	0	2	2	0
K5	0	0	0	2	2	0
P1	4	2	4	4	4	4
P2	1	0	0	0	2	0
SL1	4	2	4	2	4	2

A-3. Number of Drift Samplers, 1976

SITE	DATE						
	1	2	3	4	5	6	7
F1	0	2	0	3	0	0	0
BB1	0	3	0	0	0	3	0
KC1	0	3	0	1	0	0	0
P5	0	3	0	3	0	2	0
K2	0	3	0	0	0	3	0
BI1	0	3	0	2	0	3	0
D1	3	3	0	3	0	0	0
E1	3	3	0	3	0	3	0
SL3	0	0	0	2	0	2	0
K8	3	3	0	3	0	3	0
K1	3	3	0	3	0	3	0
SR1	3	3	0	3	0	3	0
SR3	0	3	0	3	0	2	0
K5	0	3	0	0	0	3	3
P1	3	3	0	3	0	3	3
P2	0	3	0	3	0	0	0
SL1	3	3	0	3	0	3	0
EM1A	0	0	4	0	2	3	0
EM1	0	0	6	0	4	3	0
EM2	0	0	2	0	0	0	0
EM3	0	0	5	0	4	3	0

A-4. Number of Drift Samplers, 1977

SITE	DATE 8	9	10	11	12	13	14
D3	2	0	0	0	0	1	0
N1	1	0	0	0	0	2	0
SP1	0	0	3	3	0	3	3
LI3	0	0	3	3	0	3	3
F2	1	0	0	0	0	0	0
KC2	1	0	0	0	0	0	0
SH1	3	0	0	0	0	3	0
T1	3	0	0	0	0	3	0
SG1	2	0	0	0	0	0	0
CY1	2	0	0	0	0	2	0
NR1	1	0	0	0	0	2	0
SE2	2	0	0	0	0	3	0
LI1	0	0	3	3	0	3	3
LI2	0	0	3	3	0	3	3
F1	0	0	0	0	0	1	0
KC1	2	0	0	0	0	1	0
BC1	2	0	0	0	0	3	0
P5	3	0	3	0	0	0	0
NW1	3	0	0	0	0	3	0
K2	3	0	3	0	0	3	0
BI1	3	3	2	0	0	3	0
D1	3	3	2	0	3	2	0
DC1	3	0	0	0	0	2	0
SC1	2	0	3	3	0	3	3
SE1	0	0	1	3	0	3	3
E1	3	3	3	0	3	3	0
SL2	3	0	3	0	0	3	0
K8	3	3	3	0	3	3	0
K1	3	3	3	0	3	3	0
SR2	3	3	3	3	3	3	3
SR2LA	0	0	3	0	0	0	0
SR3	3	0	3	0	0	3	0
K5	3	0	3	0	0	3	0
P1	3	3	3	0	3	3	0
P2	3	0	3	0	0	3	0
SL1	3	3	3	3	3	3	3
SL1LA	0	0	3	0	0	0	0

Number of Hester-Dendy Samplers, 1976

SITE	DATE				
	2	3	4	5	6
F1	2	0	0	0	3
BB1	3	0	0	0	3
KC1	3	0	0	0	3
P5	3	0	0	0	3
K2	3	0	0	0	3
BI1	3	0	0	0	3
D1	3	0	3	0	3
E1	3	0	3	0	3
SL3	3	0	0	0	3
K8	3	0	0	0	3
K1	3	0	3	0	0
SR1	3	0	1	0	1
SR3	3	0	0	0	3
K5	3	0	0	0	3
P1	3	0	3	0	3
P2	2	0	0	0	3
SL1	3	0	3	0	3
EM1A	0	0	0	6	6
EM1	0	6	0	6	6
EM3	0	0	0	6	6

A-6. Number of Hester-Dendy Samplers, 1977

SITE	DATE		
	10	12	14
P5	3	2	3
K2	0	0	3
D1	4	4	3
[illegible]	4	4	0
SL2	3	0	0
SR2	4	0	0
SR3	0	0	3
K5	0	0	3
P1	4	4	4
P2	3	0	0
SL1	6	4	4

APPENDIX B

Use of Study data to estimate sample sizes for a paired-site monitoring program.

The general form of the analysis of variance models for the paired-site approach has been discussed in the text of this report. Ideally, estimation of sample sizes required would be based on a pilot study or a study at a similar facility including both preoperational and operational data. This is rarely possible. However, one can use the Study data and a model including at least some of the factors involved in the true situation to estimate the error variance for the actual situation.

Sample Size

To get an estimate S, the residual standard deviation, one can proceed as follows. Drift data from two pairs of sites, P-1, SL-1 and K-1, K-8 were used because these pairs appeared similar in water quality and had been sampled more frequently than most other sites. In line with the suggestion of McKenzie et al. (1977) that analyses focus on general groups of organisms, the variable considered was the log of the ratio of total drift densities. The (geometric) mean densities for all four sites and the log of the ratio for pairs of sites are shown in Table B-3.

The simplest model for these data, using the same ratios as described earlier, is:

$$Y_{jk} = \mu + M_j + P_k + E_{jk}$$

μ = overall mean

P_k = Site pair effect

M_j = sampling time effect

Y_{jk} = The log ratio for site pair k at sampling time j.

k = 1, 2, for 2 sites

j = 1, 9 for 9 sampling periods

Table B-1. First Analysis of Variance Table for Log Ratio of Drift Densities, Site Pairs P-1, SL-1 and K-8, K-1.

Source	SS	df	MS
Site Pair (P)	.297	1	.297
Sampling Time (M)	5.16	8	.646
Residual	14.98	8	1.873
Total	20.44	17	

This analysis of variance (Table B-1) yields the estimate,

$S^2 = 1.87$ or

$S = 1.37$ log units.

However, one can also follow McKenzie et al. in considering years as replicates. This requires matching up comparable sampling periods in 1976 and 1977, and gives us data collected at six times per year over two years, with three missing dates. Then one can use the model

$$Y_{jkl} = \mu + M_j + P_k + MP_{jk} + E_{jkl}$$

M_j = sampling time effect $\quad j = 1, 6$ for 6 samples/year

$l = 1, 2$ years

Table B-2. Second Analysis of Variance Table for Log Ratio of drift Densities, site pairs P-1, SL-1 and K-8, K-1, alternative model.

Source	SS	df	MS
Site Pair (P)	2.34	1	2.340
Sampling Time (M)	2.65	5	.531
P x M Interaction	20.51	5	4.102
Residual	6.96	6	1.162
Total	32.47	17	

Note: These analyses were carried out using the program IVAN from the University of Minnesota, Department of Applied Statistics.

This analysis yields the estimates $S^2 = 1.16$, $S = 1.08$

Table B-3. Sample drift data.

Study Sample Period	j Time Index	j Season Index	Yr.	Geometric Mean Densities				Ln Ratio	
				P-1	SL-1	K-8	K-1	P-1/SL-1	K-8/K-1
1	1	2	'76	239.9	1124.	1729.	2496.	-1.544	- .3670
2	2	3	'76	198.4	1081.	870.7	1193.	-1.696	- .3151
4	3	5	'76	481.6	1991.	313.8	301.1	-1.419	+ .0411
6	4	6	'76	673.7	996.4	1263	636.0	- .3914	+..8663
8	5	1	'77	1639.	574.2	184.6	2473.	+1.1594	-2.5949
9	6	2	'77	497.4	502.0	411.0	212.5	- .009	+ .6598
10	7	3	'77	509.3	367.0	1290.	971.0	+ .3275	+ .2848
12	8	4	'77	473.9	104.1	607.8	1633.	+1.515	- .9885
13	9	5	'77	182.6	78.43	1323.	3364.	+ .8448	-..9329

Exact calculations of the power of the test given a number of samples is impossible without estimates of the variances involved in the true situation, using preoperational and operational data. However, one can estimate the sample sizes needed to compare two treatments (preoperation versus operation) using our best estimate of the variance.

Using tables (Cooke and Larntz, 1973) for the sample sizes needed to give a test with a given power, for $\alpha = .05$ and $\alpha = .01$, it is clear that the sample sizes required will be large enough (for differences of less than one log unit and the estimated variance) that use of the normal approximations will give reasonable estimates. Using the normal approximation, to detect either positive or negative changes, the formula relating the total number of sample points in both the treatment groups, n, and the α and β levels, is:

$$n = \frac{[\, Z_{1-\alpha/2} \sqrt{2\sigma_1^2 + 2\sigma_2^2} + Z_{1-\beta} \sqrt{2\sigma_1^2 + 2_2^2}\,]^2}{|d|^2}$$

where d is the magnitude of the difference to be detected. In these examples, the variances for the two groups are assumed to be equal, so that the formula reduces to:

$$t = 2 \left(\frac{s}{|d|}\right)^2 (Z_{1-\alpha/2} + Z_{1-\beta})^2$$

where $t = \frac{1}{2}n$ = the number of preoperational or postoperational data points. Then with $\alpha = .10$, $\beta = .20$, $Z_{1-\alpha/2} = 1.645$, $Z_{\beta} = .84$, and

$$t = 12.35\, S^2/|d|^2, \quad d = 3.514 \frac{S}{\sqrt{t}}$$

Thus with S = 1.077, to detect a change of .40 log units, 89 samples per treatment are necessary, which could be achieved with 3 years of preoperational data from 6 pairs of sites, 5 samples per year.

For other values of d, the number of samples, t, can be calculated by

simple substitution.

The formulas for converting a detectable difference D, pre-operational-operational) in log units, in the (ln C/T) ratio, to a percentage change are based on the formula:

$$\% \text{ change} = \frac{T_{obs} - T_{pred}}{T_{pred}}$$

where:

T_{obs} = observed geometric mean for treatment station during operation

T_{pred} = predicted geometric mean for the treatment station (predicted from the preoperational c ontrol-treatment difference and the control station operational mean).

T_{pred} is the expected mean at the treatment station when the null hypothesis (zero effect of the operation) is true, and thus ln(C/T) is expected to be constant between preoperational and operational periods. This formula for percentage change is equivalent to:

$$\% \text{ increase} = (e^{D} - 1) \times 100,$$
$$\% \text{ decrease} = (1 - e^{-D}) \times 100.$$

(McKenzie et al. 1977).

An example can clarify this equivalence.

Consider the following <u>preoperational</u> data (Table 1), where subscript g indicates the geometric mean, and superscript bars denote means.

Table 1.

$\overline{\ln C - \ln T}$	0.1
$\overline{(C/T)_g}$	1.105
$\overline{\ln C}$	4.0
$\overline{\ln T}$	3.9
$\overline{C_g}$	54.6
$\overline{T_g}$	49.4

If $\overline{\ln C}$ is assumed to be equal to 4.1 operationally, then it is possible to look

at the observed change at the treated station if the difference in $\overline{\ln C - \ln T}$ is equal to $\pm$ .4 log units, and calculate the % change in $\overline{Tg}$.

The table below presents the calculations for both cases.

	Preoperational	Case A, D=-.4 Operational	Case B, D=+.4 Operational	D=0 EXPECTED Operational
$\overline{\ln C - \ln T}$	0.1	0.5	-0.3	
$\overline{(C/T)_g}$	1.105	1.491	.7408	
$\overline{\ln C}$	4.0	4.1	4.1	
$\overline{\ln T}$	3.9	3.6	4.4	4.0
$\overline{Cg}$	54.6	52.0	52.0	
$\overline{Tg}$	49.4	36.6	81.45	54.6
% change in $\overline{T}$.		-33%	+49%	

For both cases, the expected operational $\overline{T}$ = $\overline{\ln C}$ operational - $\overline{\ln C - \ln T}$ preoperational = 4.1 - (0.1) = 4.0

hence predicted $\overline{Tg}$ = 54.6

Then the percentage change = $\frac{36.5-54.6}{54.6} \times 100 = -33\% = -(1-e^{-.4}) \times 100.$

or $\frac{81.45-54.6}{54.6} \times 100 = 49\% = (e^{.4}-1) \times 100.$