United States
Environmental Protection
Agency

EPA-600/3-81-011
April 1981

EPA

Research and Development

Lake Data Analysis and Nutrient Budget Modeling

Prepared for

Office of Water Regulations and Standards
Criteria and Standards Division

Prepared by

Environmental Research Laboratory
Corvallis OR 97330

EPA-600/3-81-011
April 1981

LAKE DATA ANALYSIS AND NUTRIENT BUDGET MODELING

By

Kenneth H. Reckhow
School of Forestry and Environmental Studies
Duke University
Durham, North Carolina 27706

Project Officer

Spencer A. Peterson
Freshwater Division
Corvallis Environmental Research Laboratory
Corvallis, Oregon 97330

CORVALLIS ENVIRONMENTAL RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
CORVALLIS, OREGON 97330

DISCLAIMER

This report has been reviewed by the Corvallis Environmental Research Laboratory, U.S. Environmental Protection Agency and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

ABSTRACT

Several quantitative methods that may be useful for lake trophic quality management planning are discussed and illustrated. An emphasis is placed on scientific methods in research, data analysis, and modeling. Proper use of statistical methods is also stressed, along with considerations of uncertainty in data analysis and modeling.

Following an introductory discussion of scientific methods, limnological variables important to lake quality management are reviewed. Methods of data acquisition, or sampling design are then presented, along with techniques for analyzing, summarizing, and presenting data (with an emphasis on robust methods). The concept of summary statistics forms a logical introduction to the next section on lake water quality indices. This is followed by methods for acquiring nutrient budget data which are of prime importance to the succeeding section on lake trophic quality modeling. Included in this section is a step-by-step procedure for the prediction of phosphorus concentration, and the estimation of the prediction uncertainty, from land use information and certain lake characteristics. At the end, some thoughts are offered on the use and limitations of the methods presented herein for lake trophic quality management planning.

CONTENTS

ACKNOWLEDGMENTS

A number of people assisted in the development and preparation of this document. Appreciation is extended to Dennis Cooke of Kent State University, and to Phil Larsen and Spencer Peterson of CERL, for providing the opportunity to prepare this report. Thanks are also due to Janine Niemer and Sue Watt for typing the document, to Paul Schneider for graphics work, and to David Lee, Ralph Ancil, Michael Beaulac, and Robert Montgomery for editorial assistance and proofreading. This document was prepared while the author was a member of the faculty in the Department of Resource Development at Michigan State University.

1. Introduction

Many useful quantitative methods exist that can be of assistance in lake quality management. Most of these methods fall under the general heading of "statistics" or "mathematical models." In this document we present some techniques from each area, but our emphasis is on methods that are applicable under the very realistic conditions of limited financial resources available for planning and of non-normal distributions of data. The methods presented are empirical, and nonparametric, or robust, whenever possible. Procedures that require few assumptions, and/or that are carried out with little investment of time and money, are stressed.

Of course, it is critical to recognize that there are often trade-offs between cost of analysis and risk associated with the resultant management decision. This is illustrated when we consider two extremes:

1. No analysis is undertaken and a decision is made based upon intuition.

2. A complete analysis is made so that outcomes associated with management options are known with certainty.

In between lie virtually all planning and management exercises. Thus, the cost of data acquisition, data analysis, and modeling must be justified in terms of benefit to the planning process. This means that our previously expressed desire for simple, low cost, methods of analysis must be tempered by the needs of the particular problem at hand.

This brief discussion of cost versus risk underscores the responsibility of the modeler, or data analyst, to the planning process. Since it is unreasonable to assume that planners are familiar with all of the tools of the modeler/analyst, the planner must, to some degree, accept the modeler/analyst's statements concerning reliability and utility of results. For that reason, quantitative analyses, or more generally, scientific research, should proceed according to some well established rules. When followed, these rules, collectively called the scientific method (Ackoff, 1962), insure that scientific studies yield credible, reliable results.

While it is not the purpose of this presentation to discuss the scientific method at length (see Reckhow and Chapra, 1980, Chapter 1), some thoughts are presented below. These represent scientific method issues that the author has found to be of concern in lake data analysis and modeling.

1. Definition: Many terms are used in limnological studies that, in part because of everyday usage, are vaguely defined. Planning depends upon useful, valuable information, and information value is a function of error. Since error can result from uncertainty in models and data, as well as from faulty communication due to confusing terminology, it is important that definitions be frequently provided. For example, what is average lake phosphorus concentration? The answer to that question depends upon the location statistic employed (see Section 3), the methods (sampling design) used to acquire the data, and the phosphorus chemical

fraction (total, ortho, ...) of concern. These should be specified when a vague term like "average" is used. As another example, consider the term, "phosphorus loading." Black box lake modelers have considered this term to mean annual areal phosphorus mass input to a lake. However, in the absence of this definition, statements made about "lake sensitivity to phosphorus loading" may be confusing or misleading (see Figure 1-1 in Reckhow and Chapra, 1980).

2. Assumptions: Implicit in all mathematical models and many statistics summarizing data are assumptions about the behavior of the system described or about limitations in the particular method or statistic employed. In order that the planner may properly weigh the quantitative information provided, the modeler/analyst should clearly specify all relevant assumptions necessary for the study conducted. For example, application of certain statistical tests is based upon an assumption of normality (see Section 3). When conducting these tests, the data analyst should identify the required assumptions and document tests for compliance (and discuss the implications of violations, if necessary).

3. Uncertainty: Uncertainty is present in all model studies because of errors in the model, in the parameters, and in the variables. Uncertainty is also present in most statistical analyses because of variability and bias. As we suggested above, uncertainty may also be introduced into an analysis because of poor communication. Uncertainty is a good measure of the value of information; as uncertainty is reduced, information becomes more precise, and hence more useful or valuable. The modeler/analyst can greatly assist the planner by specifying, whenever possible, the uncertainty in results. The planner may then use this estimate of uncertainty as indicative of the value of these results to the planning process.

4. Representativeness: In the absence of a complete census, statistics selected or calculated to represent some attribute of a system may be variable or biased. It may seem all too obvious that representativeness should be a criterion for the selection of a statistic. However, convention often interferes. For example, it is common to represent the center and spread for a data set by the mean and standard deviation. However, many "real" data sets in limnology are non-normal and highly skewed. When this situation occurs, the normal-mean-standard deviation conventional statistics are less representative than certain robust statistics (see Section 3). Often, one may face a trade-off between representativeness and some other issue, like cost of analysis. For example, in Section 5, we discuss nutrient budgets, and compare direct sampling versus nutrient export coefficients as sources of nutrient loading estimates. Export coefficients are less costly to acquire but probably less representative than the alternative. This choice, involving cost and risk, must be made according to the merits of the issue of concern. However, the modeler/analyst should consider representativeness when selecting statistics (or designing sampling programs), and he/she should justify representativeness if statistics may be in question.

5. Causality: If models and other quantitative analyses are to be useful, there must be a causal linkage between decision variables and control variables. From an understanding of theory and through sensitivity testing of the model, cause-effect relationships may be established. Without corroboration of causality, one cannot assert with confidence that selected management strategies will have the desired effect.

6. Appropriate Variable(s): There are two considerations when we think about the appropriate variable(s). First, the variable(s) for which information is gathered must coincide (or be causally-linked) with the variable(s) that impute value to the water body. Second, when a model is employed, the model variables and the decision and control variables may not always be the same. If this occurs, the modeler should strive to modify his/her analysis so that the variables of concern are included. Otherwise, the modeling will be incomplete and errors associated with a decision may be underestimated (see Reckhow and Chapra, 1980).

7. Corroboration: Models must be tested before they are applied, and this testing process has traditionally been called validation or verification. However, those terms imply truth, an attribute that a mathematical model can never achieve. Therefore, the term, corroboration (Popper, 1968), is adopted instead. Popper states that a model is corroborated when it has passed rigorous independent tests. A model that is useful for water quality management planning must be able to predict changes in water quality associated with changes in input conditions. A planning model must be adaptable. Therefore a candidate model must first be tested under conditions different from those used to calibrate the model, and a statistical goodness of fit criterion should be applied to assess the degree of corroboration. The modeler has this responsibility and the model user or planner should request documentation of these tests. Without this, there is no assurance that the model can be depended upon for accurate predictions under new conditions.

8. Cost/Risk: To briefly re-iterate an important issue in policy analysis previously stated, quantitative analyses and planning studies are not without cost (in money and time). This cost is justified only if the perceived benefit (or correspondingly, the perceived reduction in risk) from the information obtained outweighs the cost. This decision to undertake certain analyses also has a dimension of degree or thoroughness. As an analysis becomes more thorough, presumably it becomes more precise. Eventually, however, the increased level of precision may not justify the cost necessary to achieve it. This should be considered in selecting and designing planning studies.

In this brief treatment of scientific method issues, we have made some rather strong demands of modelers and data analysts in the documentation of their work. Unfortunately, some of these requirements must be tempered by the limitations in the state of the art. For example, it may not be possible to accurately assess the trade-off between cost of analysis and risk in decision making. However, the concept still holds. As long as the planner or policy analyst realizes this trade-off is part (either explicitly or implicitly) of the design of policy studies, then he/she may at least intuitively consider

the trade-off. In conducting water quality management planning, we should strive toward the conduct of analyses according to the scientific method. When this is not possible, an understanding of the concepts of the scientific method can still aid the planner by serving as an "ideal" against which to evaluate scientific studies.

One of the eight issues listed above that we should address immediately concerns the variable(s) to be studied. While this issue is problem-specific, the U.S. Environmental Protection Agency (Larsen, 1980) has developed a list of variables that is likely to contain the variables of concern in most lakes. This list, presented in Exhibit 1, is broken into two parts. The methods presented in the remainder of this chapter are more applicable to the "General Lake Quality" variables in part A of Exhibit 1. In particular, the problem of eutrophication is emphasized herein, so techniques oriented to the study of trophic variables predominate. However, many of the methods can be useful for other limnological variables. In addition, an effort has been made to stress concepts, so that the reader may understand scientific and statistical inference independent of the direct utility of the specific techniques.

Exhibit 1. Liminological variables of importance in lake management (Larsen, 1980).

A. General Lake Quality Variables

- phosphorus
- nitrogen
- dissolved oxygen
- turbidity (Secchi disk)
- chlorophyll a
- macrophyte coverage
- bacteria and viruses
- toxic substances

B. Use-Specific Variables

1. Swimming	2. Fishing	3. Boating
temperature (air/water)	planting/stocking programs	macrophytes
turbidity	fish type and abundance	algae
algal abundance	dissolved oxygen	obstructions
macrophytes	toxic substances	trash
odor (dissolved oxygen)	algae	facilities
disease-causing organisms	macrophytes	lake size/ depth
parasites and insects	spawning grounds	
toxic substances	temperature	
oil		
trash		
facilities		
beach and bottom type		

2. Acquisition of Lake Data

Lake data are acquired because there is a need for the data. For lake quality management, this need is reflected in the value of the information provided by the data. The purpose, therefore, of this section is to provide guidance in the establishment of cost-effective data gathering programs.

Although much of this section is devoted to statistical sampling design, "data acquisition" is purposely used in the section title to underscore the notion that data may be obtained by means other than sampling. For example, many limnological issues may be completely or partially addressed using existing data. Alternatively, existing data on surrogate variables may prove useful after statistical analysis is used to quantify the relationship between the surrogate and the quality variable of concern. As we have stressed in the last section, however, the decision to use existing data must be made with some understanding of the cost/risk trade-offs. Acquisition of existing data is almost always less costly than sampling to obtain new data. However, existing data may be less representative of the issue of concern than new data, and this non-representativeness translates into greater risk in decision making. The planner must consider these trade-offs when designing data acquisition programs.

Most likely, some or all of the data needed for lake quality management planning will be obtained under a sampling program that should be designed using statistical methods. Before we survey these methods, it is instructive to discuss some concepts inherent in statistical sampling design. Consider the words used to identify this topic: "statistical sampling design." This task is called "sampling" because only a limited amount of information is obtained. The entirety of the characteristic sampled is called the population. Statistics obtained through sampling are called sample statistics and they are intended to represent the population, or true, values. Sampling is undertaken because it is often infeasible to survey the total population. For example, it is clearly impossible to survey an entire lake throughout time and space for a population value for algal biomass. Instead we turn to sampling and undertake a program to obtain a representative sample statistic. This is where the other terms in "statistical sampling design" become important. Sampling is a problem in "statistical design" because statistical methods help us design a program that yields representative data.

Statistical sampling design, has, as a basic consideration, the trade-off between uncertainty and cost. Uncertainty results from variability, error, and bias. Variability exists because of natural fluctuations inherent in a characteristic (e.g., natural variations in stream or lake phosphorus concentration), or because of uncertainty inherent in a statistic used to summarize a set of data. Errors may arise in any of the individual steps of sampling, measurement, analysis, and estimation. Bias may result from a number of causes, all associated with the fact that a sample may not be representative of the population from which it was drawn. For example, a survey of a stratified lake consisting of fifty concentration samples, with only one taken from the hypolimnion, probably will yield a biased statistic for mean concentration.

When sampling programs are designed, variability, error, and bias should be estimated for all candidate designs. In this manner, the trade-off between uncertainty and cost of sampling can be as explicit as possible. The trade-off can be evaluated in terms of financial constraints and needs for data reliability for the selection of an appropriate design.

In order to understand the statistical relationships that are used to design sampling programs, there are some statistical terms that must first be defined. The terms result from expected value theory and are most useful with well-behaved symmetric probability density functions, like the normal distribution.

1. Mean: The mean is a measure of location of central tendency for a distribution or set of data. The mean, $\bar{x}$, is

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad (1)$$

where: x_i = data point i
n = total number of data points.

For symmetric distributions of data, the mean is a reliable statistic to use to represent the average or central tendency. However, it must be emphasized that, when representing a set of data with descriptive statistics, our true objective is to select a statistic that best indicates the distribution center, and not simply to calculate the mean. Sometimes the mean is the appropriate statistic, and sometimes it is not. Other candidate statistics for location include the median, mode, geometric mean, trimmed mean, tri-mean, and biweight (see Reckhow and Chapra, 1980, or Mosteller and Tukey, 1977 for discussion and analysis). Some of these are presented in the next section.

2. Variance and Standard Deviation: The variance and standard deviation are measures of spread or scale for a distribution or set of data. The variance, s^2, is:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} \qquad (2)$$

The standard deviation, s, is simply the square root of the variance:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}} \qquad (3)$$

Here, too, we should recognize that the standard deviation or variance is not, by definition, the spread in a distribution. Rather it is a statistic chosen to represent spread. For certain symmetric distributions, the standard deviation is a good choice. For other distributions, notably skewed distributions, alternative measures of spread, such as the average deviation, the range, or the interquartile range may be more appropriate (see Reckhow and Chapra, 1980, and Mosteller and Tukey, 1977). Some of these alternatives are presented in the next section.

3. Standard Error of the Estimate: The standard error of the estimate is the mean square error for a statistic or estimate. Often the statistic of concern is the mean, so we use the symbol, $s_{\bar{x}}$, and calculate the standard error as:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad (4)$$

The standard error of the estimate is a measure of precision of a statistic. To the extent that precision and uncertainty are equivalent, $s_{\bar{x}}$ is also a measure of uncertainty. However, we mentioned earlier that uncertainty includes variability and bias. In the absence of supplemental uncertainty[1] (see Reckhow and Chapra, 1980 and Mosteller and Tukey, 1977), precision accounts for variability but not bias. Therefore, while we use the standard error of the estimate extensively in sampling design, we must be wary of its limitations. $s_{\bar{x}}$ is a measure of variability in a statistic, such as the mean. This may not be equivalent to the uncertainty in central tendency for a population. Generally, our true concern is with the latter, not with the former.

[1] Supplemental Uncertainty is uncertainty that is not measured by the statistic employed (in this case, the standard error of the estimate). For example, supplemental uncertainty exists when data are not truly representative of a characteristic. Since $s_{\bar{x}}$ is data-derived, there must be additional uncertainty associated with nonrepresentativeness.

4. Coefficient of Variation: The coefficient of variation, cv, is:

$$cv = \frac{s}{\bar{x}} \tag{5}$$

This statistic is a useful measure of relative variability. It is a dimensionless quantity that facilitates comparison among dispersion statistics by expressing the standard deviation as a fraction of the mean.

The design of a sampling program is often expressed in terms of random sampling. In theory, random sampling refers to data acquisition when individual points are selected by chance. Under random sampling, all members of a population are equally likely to be chosen in the sample. In practice, however, limnological sampling is rarely random. It is usually systematic in space (i.e., sampling occurs at pre-specified sites) and systematic, or systematic with a random start (i.e., begun on a randomly chosen day and continued on systematically pre-specified days thereafter) in time. Statistical relationships used in the design of sampling programs are generally aimed toward random sampling or a variation thereof. However, with an understanding of limnological relationships, the rudiments of sampling design, and possible sources of supplemental uncertainty, we can often apply random sampling design relationships to the systematic sampling programs that we often adopt in limnology. In particular, random sampling design equations may be used for systematic sampling if there is no bias introduced by incomplete design, and if there is no periodic variation in the population measured. Use is further justified if the systematic sampling begins with a random start.

There are certain quantities that are common to most sampling design relationships. These include the number of samples, the desired precision (or error) of the estimate, and the inherent variability in the characteristic measured. The quantities are all present in the relationship for the standard error of the estimate, Equation 4. When we invoke the common assumption that the data are normally distributed,[1] we can use the t-statistic (see the next section) to specify the confidence level desired in our sample. Thus, for simple random sampling, Equation 4 is modified to yield:

$$n = \frac{t^2 s^2}{d^2} \tag{6}$$

---

[1] The central limit theorem states that the distribution of $\bar{x}$ for sufficiently large samples and for any population with a finite variance, will be normal. ("Sufficiently large" is determined by the degree of normality of the population and the acceptable error; 30 to 100 samples may be required depending upon these issues (Blalock, 1972).) This justifies the use of the t-statistic with the standard error of the mean. However, when the distribution of concern is severely non-normal, robust statistics (see Section 3) should be employed, and sampling design may be conducted on a somewhat ad hoc basis.

where: n = number of samples
t = student's t-statistic
s^2 = population variance estimate
d = desired precision.

Equation 6 may be used to estimate, for random (or "effectively" random) sampling, the number of samples necessary to achieve a desired level of precision, given an estimate of population variability. Desired precision is selected after consideration of the acceptable error, the inherent variability of the characteristic sampled, and the sampling cost. The sampling design decision can be expressed as a trade-off between desired precision and cost, if the number of samples is re-expressed in terms of sample cost. For example, one common cost function is:

$$C(n) = c_0 + c_1 n \tag{7}$$

where: $C(n)$ = total cost of sampling
c_0 = initial fixed cost
c_1 = cost per sample.

When Equations 6 and 7 are combined, a random sampling design may be specified according to either desired precision or a cost constraint.

Now, to use Equations 6 and 7 for sampling design, an estimate of the population variance is needed. In theory, we want to estimate the variance using Equation 2 on normal-like data. In practice, we are really interested in the "vague concept" (Mosteller and Tukey, 1977) distribution spread, which may or may not be best estimated by a sample variance. Further, we rarely have sufficient data on the characteristic to be sampled to reliably calculate a variance. (If we did, this might call into question the decision to sample.) Therefore, we must depend upon a variety of methods for a measure of distribution spread that can be used in Equation 6 for population variance. These methods include (Cochran, 1963):

1. Use existing information on the population to be sampled, or existing information on a similar population.

2. Rely on informed judgment, or on an educated guess.

3. Undertake a two-step sampling procedure. Use the results from the first step to estimate the required terms in Equation 6. Then use this to design the second step. Data from both steps may be employed in the final estimate of the characteristic of interest.

4. Conduct a pilot study on a convenient or particularly meaningful subsample. Use the results to estimate the required terms in Equation 6. Unlike for two-step sampling, the pilot study results are generally not used in the final estimate of the characteristic of interest. This happens because the pilot sample is often not representative of the population as a whole. This possible non-

representativeness must be taken into account when the pilot survey results are used to estimate variance. A modification might be necessary if it is thought that the pilot survey provided an overestimate or underestimate of the population values.

For many types of problems, sampling can be more efficient when the design is based on the fact that a population often contains strata that are homogeneous within and heterogeneous with respect to other strata. For example, stratified lakes generally exhibit homogeneous conditions within the epilimnion and the hypolimnion, while at the same time these two strata are heterogeneous with respect to each other. As another example, the nutrient flux to a lake can vary significantly from tributary to tributary. In these situations, sampling is more efficient when sample numbers are allocated according to stratified random sampling design. Then within each stratum, sampling is random or systematic with a random start. Sampling is allocated in stratified random sampling design according to:

$$\frac{n_i}{n} = \frac{w_i\ s_i}{\Sigma(w_i\ s_i)} \tag{8}$$

where:
- n_i = number of samples in stratum i
- n = total number of samples
- w_i = a weight reflecting the size (number of units, for example) of stratum i
- s_i = standard deviation of sampled characteristic within stratum i.

If sampling cost may be estimated by:

$$c = c_o + \Sigma(c_i\ n_i) \tag{9}$$

then:

$$\frac{n_i}{n} = \frac{w_i\ s_i/\sqrt{c_i}}{\Sigma(w_i\ s_i/\sqrt{c_i})} \tag{10}$$

In order to apply Equation 8 or 10, a relationship is needed for the total number of samples, n. Two equations are available, depending upon whether precision or cost if fixed beforehand. If precision is fixed (at d), and cost may be estimated according to Equation 9, then (Cochran, 1963):

$$n = \frac{\left(\Sigma w_i\ s_i\ \sqrt{c_i}\right)\Sigma\left(w_i\ s_i/\sqrt{c_i}\right)}{d^2/t^2} \tag{11}$$

If cost is fixed, then (Cochran, 1963):

$$n = \frac{(c - c_o)\ \Sigma\ (w_i\ s_i/\sqrt{c_i})}{\Sigma\left(w_i\ s_i\ \sqrt{c_i}\right)} \qquad (12)$$

In summary, the composition of the stratified random sampling design equations leads to the following general conclusions concerning stratified sampling. A larger sample should be taken in a stratum if the stratum is:

1. more variable (s)
2. larger (w)
3. less costly to sample (C).

Example 1

To illustrate how samples that are acquired without concern for statistical design may be quite misleading, a hypothetical example is constructed. For ease of explanation, assume that Exhibit 2 is a complete description of the population of phosphorus concentration values (in micrograms per liter) in a stratified lake. The values in the exhibit were randomly generated from three lognormal distributions.[1] Using μ and σ as symbols for the population mean and standard deviation, the distribution parameters are:

	Epilimnion	Metalimnion	Hypolimnion
μ (log transform)	1.301	1.544	1.700
μ (μg/l)	20	35	50
σ (log transform)	.146	.089	.093
$\mu + \sigma$ (μg/l)	28	43	62
Number of cells	61	35	41

The "population" statistics that best represent the center of the population of lake phosphorus concentration values are probably the weighted geometric mean[2] and the median. These statistics are:

weighted geometric mean = 30.4 μg/l
median = 34 μg/l

---

[1] After the values were generated, they were placed in the lake diagram in order to best approximate realistic concentration contours and gradients.

[2] The geometric mean is the antilog of the mean of a lognormal distribution. In this example, the geometric mean of each stratum is weighted according to the stratum's percentage of cells, for the calculation of the weighted geometric mean.

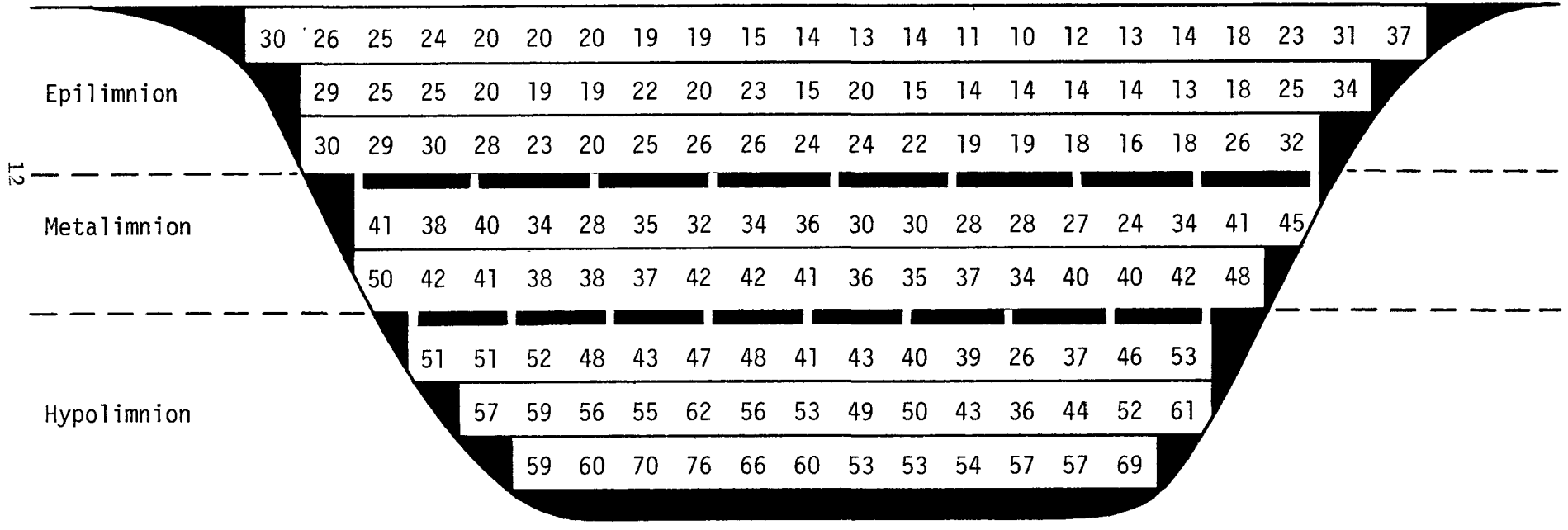

Exhibit 2. Lake phosphorus concentration (µg/l) for Examples 1 and 2.

The results of sampling should be compared to these statistics for a measure of the success of the sampling program.

Now, suppose that we undertake a brief sampling program in order to estimate the average phosphorus concentration in the lake. Consider the following examples illustrating how this might be done.

A. Take a single depth profile in a deep section of the lake. Randomly selecting three profiles, we find:

1. measurements (µg/l): 11, 14, 19, 28, 37, 39, 43, 54

mean = 30.6 µg/l
median = 32.5 µg/l

2. measurements (µg/l): 19, 20, 26, 32, 42, 47, 62, 76

mean = 40.5 µg/l
median = 37 µg/l

3. measurements (µg/l): 14, 15, 22, 30, 35, 40, 50, 53

mean = 32.4 µg/l
median = 32.5 µg/l

B. Take surface samples only. Randomly selecting eight samples we find:

measurements (µg/l): 30, 26, 24, 20, 19, 11, 14, 18

mean = 20.3 µg/l
median = 19.5 µg/l

C. Take surface-to-bottom samples at three randomly selected sites:

measurements (µg/l): 20, 20, 28, 40, 41, 51, 57, 13, 20, 24, 30, 36, 43, 49, 53, 14, 13, 18, 34, 42, 53, 61

mean = 34.5 µg/l
median = 35 µg/l

D. Take four samples from any site and depth in the lake. Randomly selecting three sampling programs, we find:

1. measurements (µg/l): 26, 50, 14, 34

mean = 31 µg/l
median = 30 µg/l

2. measurements (µg/l): 13, 18, 20, 28

mean = 19.8 µg/l
median = 19 µg/l

3. measurements (µg/l): 20, 26, 43, 61

mean = 37.5 µg/l
median = 34.5 µg/l

While we must be careful in drawing conclusions from a small sample of sampling programs, there are a few results in the examples presented above that are consistent with the findings of many lake sampling experiences.

1. Surface sampling can lead to biased estimates of average conditions in a stratified lake. Underestimation is often the result.

2. Depth profile sampling is preferred to single layer (stratum) sampling, particularly if samples taken are roughly proportional to the stratum volume.

3. A small number of samples (example D) is more apt to result in a biased estimate of average conditions than is a large number of samples (example C).

Example 2

Let us use stratified sampling design to develop a sampling program for the lake in Exhibit 2. Assume that:

1. The samples taken in Example 1C above represent existing data from which the sampling program will be designed. Because the number of samples is small, the standard deviation (s) will be estimated as one-half the range of data points within each stratum. In an actual lake sampling program, measurements could be assigned to a stratum on the basis of a temperature profile.

2. The size (w) of each stratum will be estimated by the relative number of stratum measurements in Example 1C. In an actual lake sampling program, the size of a stratum would be determined by its volume.

3. It is desired that a sampling program be designed to provide an estimate of mean phosphorus concentration that is within ± .005 mg/l of the true mean at the 95% level.

From the samples in Example 1C, we have the following breakdown.

Measurements (µg/l)

epilimnion: 20, 20, 28, 13, 20, 24, 14, 13, 18
metalimnion: 40, 41, 30, 36, 34, 42
hypolimnion: 51, 57, 43, 49, 53, 53, 61

The necessary statistics are:

	Epilimnion	Metalimnion	Hypolimnion
Range (µg/l)	15	12	18
s (µg/l)	7.5	6	9
w	9/22	6/22	7/22

To design the sampling program, first solve for the total number of samples to be taken, using Equation 11 with cost (c_i) constant across all sampling sites.

$$n = \frac{(\Sigma w_i \, s_i)^2}{d^2/t^2}$$

for the sample sizes under consideration here, t-2. Therefore:

$$n = \frac{[(9/22)\ (7.5) + (6/22)\ (6) + (7/22)\ (9)]^2}{5^2/2^2}$$

$$n = \frac{57.28}{6.25} = 9.16$$

$$n \sim 9 \text{ samples}$$

Equation 10 may be used to allocate the samples among the strata (again with cost constant across sites).

$$\frac{n_i}{n} = \frac{w_i \, s_i}{\Sigma(w_i \, s_i)}$$

for the epilimnion:

$$\frac{n_e}{n} = \frac{(9/22)\ (7.5)}{(9/22)\ (7.5) + (6/22)\ (6) + (7/22)\ (9)}$$

$$n_e = 3.71$$

for the metalimnion:

$$\frac{n_m}{n} = \frac{(6/22)\ (6)}{(9/22)\ (7.5) + (6/22)\ (6) + (7/22)\ (9)}$$

$$n_m = 1.98$$

for the hypolimnion:

$$\frac{n_h}{n} = \frac{(7/22)\ (9)}{(9/22)\ (7.5) + (6/22)\ (6) + (7/22)\ (9)}$$

$$n_h = 3.47$$

Since samples can be taken only in integer units, and given the nature of the results calculated above, we might recommend that 10 samples be taken, and they be distributed 4, 2, and 4 in the epilimnion, metalimnion and hypolimnion, respectively. As an approximate check, the following samples are chosen randomly.

epilimnion (µg/l): 30, 13, 19, 24
metalimnion (µg/l): 40, 28
hypolimnion (µg/l): 52, 41, 50, 57

From this sample, the following statistics may be calculated.

1. a volume-weighted mean

$$\bar{x}_w = \sum^{s}\left(\frac{w_s}{n_s}\sum^{i}(x_{si})\right) \qquad (13)$$

where: subscript s refers to strata
subscript i refers to samples

$$\bar{x}_w = (1/4)\ (9/22)\ (30 + 13 + 19 + 24) + (1/2)\ (6/22)\ (40 + 28)$$
$$+ (1/4)\ (7/22)\ (52 + 41 + 50 + 57)$$

$$\bar{x}_w = 34.0\ \mu g/l$$

2. a volume-weighted standard deviation, which is estimated from one-half the range within each stratum, because of the small sample size:

$$s_w^2 \sim [(9/22)\ (8.5)]^2 + [(6/22)\ (6)]^2 + [(7/22)\ (8)]^2$$

$$s_w \sim 4.6\ \mu g/l$$

3. a volume-weighted standard error:

$$s_{\bar{x}_w}^2 \sim \tfrac{1}{4}[(9/22)\ (8.5)]^2 + \tfrac{1}{2}[(6/22)\ (6)]^2 + \tfrac{1}{4}[(7/22)\ (8)]^2$$

$$s_{\bar{x}_w} \sim 2.45\ \mu g/l$$

It is shown in the next section that the precision of the estimate of the mean at the 95% level is:

$$\bar{x} \pm t_{.05}\ s_{\bar{x}}$$

For this problem, the precision is approximately:

$$\bar{x}_w \pm 2\ s_{\bar{x}_w}$$

or:

$$34.0\ \mu g/l \pm 4.9\ \mu g/l$$

The means that the 95% confidence interval for the volume-weighted mean is

$$29.1\ \mu g/l < \mu_w < 38.9\ \mu g/l$$

A couple of final observations are in order. First, note that the true median (34 µg/l) is well within the 95% confidence limits but that the true geometric mean (30.4 µg/l) is just slightly inside. Also note that both true values are within the pre-specified confidence interval (± .005 µg/l). Our actual interval at the 95% level (± .0049 µg/l) is lower than the pre-specified value because we chose to take 10 samples (versus 9 or 9.16) and because our sample turned out to be relatively homogeneous.

In concluding this section it is worthwhile to mention useful references for sampling design. Many excellent books and monographs have been written about sampling design, and the reader should consult one or more of them if additional details on this topic are desired. Among the recommended references are Cochran (1963); Hansen, Hurwitz and Madow (1953); Jessen (1978); Williams (1978); and Freese (1962). Noteworthy among these are Williams, as an introduction to sampling design, and Cochran, as a more advanced text and as an excellent reference. In addition, some statistics books contain sections on sampling design; Snedecor and Cochran (1967) is one recommended example.

3. Analysis of Lake Data

Once data have been acquired, either through a sampling program or from existing sources, it is usually necessary to summarize the data in a few well-chosen statistics to make the results most useful for planning. Traditionally, these chosen statistics are those statistics important in expected value theory and normal distribution theory (e.g., the mean and standard deviation). Often real data sets are misrepresented by these "traditional" statistics, so we adopt a different approach in this section. First, we present the "vague concept" (Mosteller and Tukey, 1977) for which a particular statistic (such as the mean, or median) is selected. Then we offer a few options for statistics to represent the vague concept, mentioning some pros and cons for each. Throughout this section, in fact, we try to present more than one option for a statistical exercise. This should foster the correct notion that use of the traditional methods should represent a choice. The other options which we call robust statistics, robust methods, or nonparametric techniques may in many instances be the superior choice, however. The material presented in this section, and the references cited, should help the reader make this choice.

The first exercise one should conduct with a set of data is to plot the data on a graph. For data on a single variable, the frequency plot or histogram is useful. A modification of the traditional bar histogram which we present here is the stem and leaf plot (Tukey, 1977). Unlike the histogram, however, the stem and leaf diagram retains the numbers (i.e., the individual data points) in the display, and their relative abundance yields the distribution shape.

Example 3

(From Reckhow and Chapra, 1980)

To illustrate an alternative to the bar histogram, let us take the data in Exhibit 3 and create two stem and leaf diagrams. A stem and leaf diagram (Tukey, 1977; Mosteller and Tukey, 1977) is constructed from a set of data with the higher digits (the "tens" and "hundreds" digits in Exhibit 3) forming the left side of a column as in Exhibit 4. On the right side of the column, the lowest ("units") digit for each data point is placed in a row opposite the

Exhibit 3. Phosphorus and Chlorophyll *a* data.

Total Phosphorus Concentration (µg/l)	Chlorophyll *a* Concentration (mg/l)
5	1.4
7	3.0
8	1.7
10	2.1
10	2.0
15	6.0
18	4.9
24	22
29	8.2
30	12
32	25
33	14
38	12
41	20
42	24
43	30
48	20
68	42
84	84
92	103
96	120

Exhibit 4. Stem and leaf diagrams.

A) Phosphorus Concentration (µg/l)		B) Chlorophyll *a* Concentration (µg/l)	
0	578	0	13222658
1	0058	1	242
2	49	2	25040
3	0238	3	0
4	1238	4	2
5		5	
6	8	6	
7		7	
8	4	8	4
9	26	9	
10		10	3
11		11	
12		12	0

appropriate higher digit. Thus, in Exhibit 4A, the entries in the 0-row represent 5, 7, and 8 µg/l of phosphorus, and the entries in the 1-row represent 10, 10, 15, and 18 µg/l of phosphorus. In Exhibit 4B, concentrations are rounded off to the nearest integer.

The advantage of a stem and leaf diagram is that it provides most of the features of a histogram while retaining the numerical values of a table of data. Like a histogram, the stem and leaf display can be constructed using different data groupings (e.g., the right-side digit could be the tens digit, or any other digit, if appropriate). However, the stem and leaf diagram is not as flexible as the histogram, in that stem and leaf diagrams are constrained to order-of-magnitude changes in groupings (e.g., histogram data can be grouped: 1-4, 5-8, 9-12, ..., whereas stem and leaf data are always grouped in some multiple of ten: 0-9, 10-19, 20-29, ...).

Another useful graphical procedure for univariate data is the box plot (Tukey, 1977; McGill et al., 1978). The box plot is constructed largely from the order statistics, and it provides information on the median, spread or variability, skew, size of data set, and statistical significance of the median. All of this information may be conveyed on a graph in essentially the same space used to plot the mean and standard deviation.

Box plots for the phosphorus data in Exhibit 5 for five lakes are drawn in Exhibit 6 with median chlorophyll a on the x-axis. To construct a box plot for a set of data on a single variable, the steps listed below may be followed.

1. Order the data from lowest to highest.

2. Plot the lowest and highest values on the graph as short horizontal lines. These represent the extreme values for each box plot, and they identify the range.

3. Determine the upper and lower quartiles (the data points at the 25 and 75 percentiles) for the data set. These values bound the interquartile range (I), which is the "distance" between quartiles. The quartiles define the upper and lower box edges, and they are connected to the respective range values.

4. Plot the median as a dashed horizontal line within the box.

5. Select a scale so that the width of the box represents the sample size, or the size of the data set used to construct each box. For example, the width of the boxes may be set as proportional to the square root of the sample size (n). Then, if n = 10 is represented by one centimeter of width, the width of all the boxes may be calculated based on their sample size.

6. Determine the height of the notch (in the box at the median) based on the statistical significance of the median. The standard deviation (s) of the median may be estimated by:

$$s = 1.25\ I/1.35n \quad (14)$$

for a range of distributions with normal-like centers (McGill et al., 1978). The height of the notch above and below the median is ± Cs:

$$\text{Notch Limits} = \text{Median} \pm Cs \quad (15)$$

Exhibit 5. Phosphorus and chlorophyll data for five lakes.

Phosphorus (µg/l)				
Lake A	Lake B	Lake C	Lake D	Lake E
5	18	180	54	115
8	28	116	23	97
11	15	176	49	84
12	37	117	20	161
15	25	118	34	116
16	13	113	52	121
7	93	115	27	174
7	47	132	20	102
7	25	125	46	91
4	20	110	22	110
6	22	115	25	88
10	50	145	44	144
11	40	140	38	153
Chorophyll a (µg/l)				
2.6	8.5	65.7	39.0	31.1
4.1	4.2	31.0	16.2	20.4
3.5	4.7	42.1	42.0	21.6
9.0	35.3	30.2	14.4	1.5
5.6	6.5	30.0	23.5	2.1
7.4	12.1	14.2	20.4	2.8
1.9	20.4	9.6	31.5	14.4
2.3	20.4	25.9	28.9	12.0
2.6	7.3	19.6	20.9	17.1
2.8	8.2	21.2	18.2	7.3
1.7	5.1	23.0	23.0	6.1
6.1	15.0	51.3	35.4	25.4
7.7	10.2	47.1	31.8	26.8

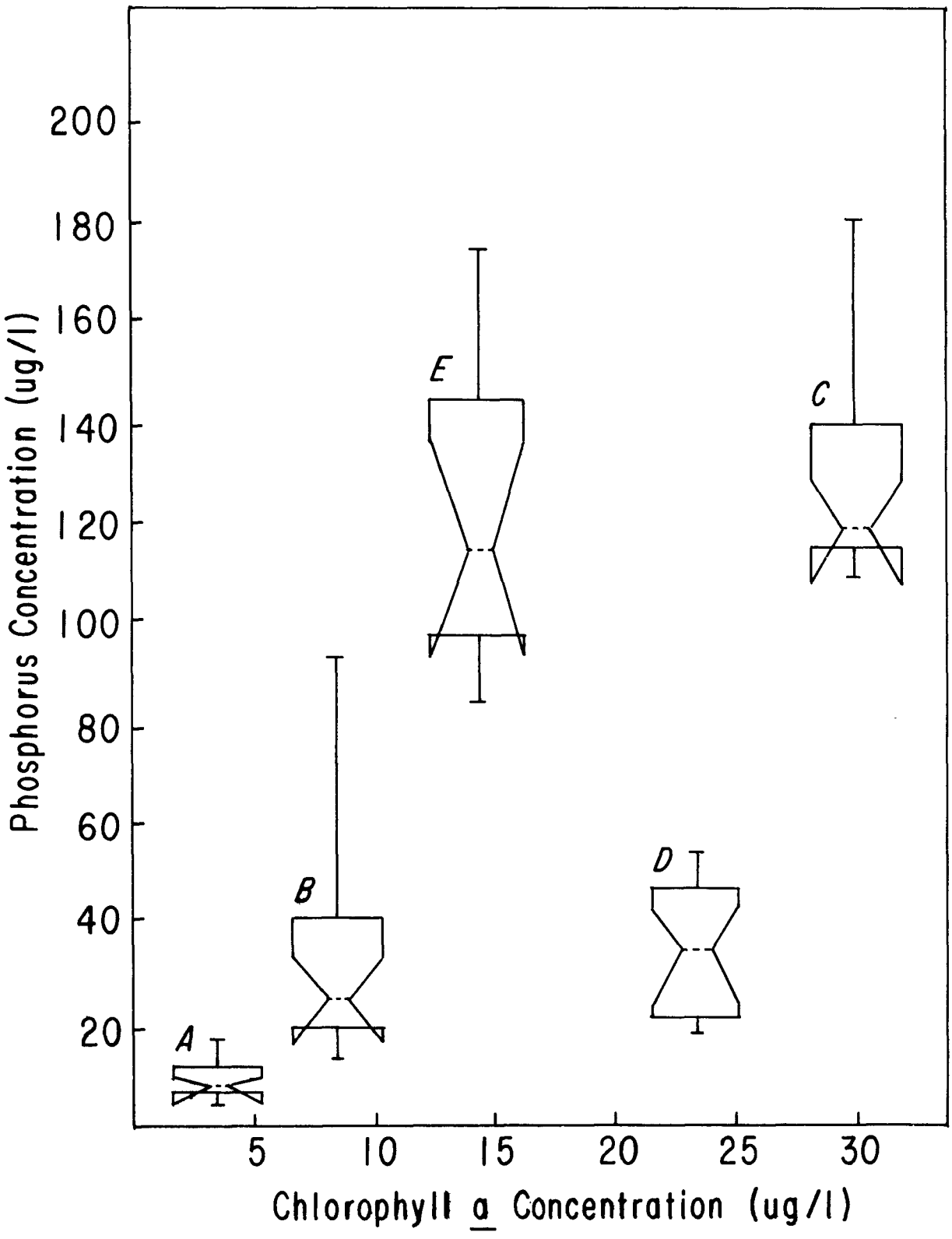

Exhibit 6. Box plots for phosphorus concentration.

C is a constant that lies between 1.96 (appropriate if the standard deviations for the data sets are quite different) and 1.39 (preferrable when the standard deviations are nearly identical). McGill et al. chose a compromise value of 1.7 for their example, and that value was also used in Exhibit 6. THus the notch heights are:

$$\text{Median} \pm 1.7\ (1.25\ I/1.35n)$$

With this mathematical definition of the notch heights, the notch in the box provides an approximate 95% confidence interval for comparison of box medians. Therefore, when the notches for any two boxes overlap in a vertical sense, these medians are not significantly different at about the 95% level.

The box plots present the following information:

1. the median
2. the interquartile range, which is a measure of spread or variability
3. the range (maximum value minus minimum value), and an impression of skew through a visual comparison of the symmetry above and below the median
4. the size of the data set, which is an indication of the robustness of the statistics
5. the statistical significance of the median.

Box plots may be used for a variety of purposes both in the display of data and in the examination of data. For example, Reckhow (1980) adds two symbols to the box plot in Exhibit 6, representing average influent phosphorus concentration and lake phosphorus concentration predicted to coincide with significant hypolimnetic oxygen depletion. Since these modifications, coupled with the box plot, probably represent a unique view of the data, new empirical insights may be likely. A second addition to the box plot proposed by Reckhow is an overlay of the prediction and prediction interval for a proposed phosphorus lake model. This might represent another form of residuals analysis in the model development process. Although the box (i.e., I) and the prediction interval do not represent the same "level" of spread statistic, a comparison of these two regions should enhance the traditional residuals comparison of two points (predicted and observed location statistics). Another use for the box plot has been recently proposed by Simpson and Reckhow (1980) in their work on discriminant analysis of algal dominance in lakes. They found the box plot extremely useful for the identification of variables that may be used to discriminate between two pre-selected groups of cases. These discriminating variables were identified by the degree of overlap of the boxes and notches, when the box plots - one for each group and variable - are compared (the greater the degree of overlap, the less discriminating the

variable). Undoubtedly other applications of the box plot will be proposed, but even in its unmodified form, the box plot should become a standard method for the presentation of data.

After data have been plotted and the shape and/or trend of the distribution of data have been ascertained from the grpah(s), it is often desirable to summarize the data in a few well-chosen statistics. These statistics should be selected to represent certain "vague concepts" (Mosteller and Tukey, 1977) concerning a set of data. The most important of the vague concepts are "central tendency" and "spread." The central tendency, or center, of a set of data can be represented by the mean, median, mode, geometric mean, and other similar location statistics. The spread of a distribution of data is indicated by the standard deviation, interquartile range, mean absolute deviation, median absolute deviation, range, and other statistics representing scale.

Since most scientists and engineers learn statistics from a basis of normal distribution theory, there is a tendency to always summarize a set of data with the mean and the standard deviation (or variance). This tendency developed because the mean and standard deviation are "sufficient statistics" for the normal distribution. In other words, the mean and the standard deviation completely describe a distribution when it is normal. Unfortunately, many sets of data representing acutal limnological characteristics exhibit highly non-normal distributions. In those situations, the vague concepts become important, and sample statistics should be chosen to represent central tendency, spread, and other relevant characteristics of the distribution.

Candidate statistics are presented below for central tendency and spread. Certain of these statistics are called "robust" because they represent the appropriate vague concept well for a variety of distribution shapes. Selection of the best statistic to quantify a particular vague concept is dependent upon the distribution of the sample data, the need for statistic robustness, and mathematical convenience. As a general rule, the normal theory statistics (mean and standard deviation) are favored in situations when sample data are roughly normal or uniform in distribution and/or when mathematical tractability is important. Robust statistics (e.g., the median and interquartile range) are generally perferred when the data describe a skewed or irregularly shaped distribution, or when insufficient information is available to characterize the shape of a distribution. See Reckhow and Chapra (1980) for additional discussion concerning the choice of appropriate statistics.

1. Measures of Central Tendency

   a. Mean, $\bar{x}$:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad (16)$$

The mean is the most commonly used location statistic. Note that since the mean is an equally-weighted sum of the observations, extreme values of x_i can have a strong influence on $\bar{x}$. For this reason, the mean is not robust under conditions of distribution skew.

b. Median. The median is the middle value in a set of data when the data are ordered from low to high. Since the median is unaffected by the particular values assumed by the ordered data points, it is robust in situations with extreme data (i.e., skewed distributions).

c. Mode. The mode is the single value most frequently observed. For a probability density function or histogram, it corresponds with the peak, or most likely value.

d. Geometric Mean. The geometric mean is equivalent to the antilog of the mean of a set of log-transformed data. This is an important statistic for many hydrologic and water quality variables that are approximately characterized by lognormal distribution. For log-normally-distributed data, the geometric mean is probably the best central tendency statistic.

2. Measures of Spread

a. Standard Deviation, s:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \tag{17}$$

Like the mean, the standard deviation is an often employed statistic. Also like the mean, the standard deviation is not robust under conditions of distribution skew. In particular, since the deviations (from the mean) are squared, data points with large deviations (outliers) have a strong impact on the magnitude of the standard deviation.

b. Interquartile Range, I. When data are ordered from low value to high value, the interquartile range is the difference between the value at the 75% level and the value at the 25% level. Since the interquartile range, like the median, is based upon order statistics, it is robust in situations with extreme data.

c. Mean Absolute Deviation and Median Absolute Deviation. The absolute deviation is defined as:

$$\text{Absolute Deviation} = \left| x_i - \bar{x} \right| \tag{18}$$

The mean absolute deviation is the mean value among the absolute deviation data points, while the median absolute deviation is the median value among the absolute deviation data points. The choice between these absolute deviation statistics is equivalent to the choice between the mean and median as summarized above.

d. Range. The range is the difference between the highest value and the lowest value. While it is an easy statistic to calculate, it is obviously sensitive to extreme data. Nevertheless, the range is an important indicator of distribution spread.

Following the selection and calculation of sample statistics, there is frequently a need to test or quantify certain relationships about the population(s) of concern. This exercise could take the form of an hypothesis test, confidence intervals, or perhaps a goodness-of-fit test. In this brief discussion, the presentation is limited to two methods for hypothesis testing. However, it is important to realize that sometimes confidence intervals are more appropriate when comparing statistics or data sets. Reckhow and Chapra (1980), Wonnacott and Wonnacott (1972), and other statistics texts (identified at the end of this section) examine the pros and cons of hypothesis testing, and suggest appropriate uses for confidence limits and hypothesis tests.

The tests presented below are the t-test, the standard statistical test associated with normal distribution theory, and the Mann-Whitney or Wilcoxon test, the most commonly applied nonparametric, or distribution-free test. For either procedure, the test is begun with the establishment of a "null" hypothesis. This null hypothesis is often proposed as a "straw man," based on a suspicion that it is false. Competing with the null hypothesis for acceptance is the alternative hypothesis. Under this scheme, then, there are four possible outcomes associated with the fundamental truth or falsity of the hypotheses, and the success or failure of the hypothesis testing.

The t-test is based on assumptions of sampling from normal distributions, homogeneity of variances, and independent errors. The Mann-Whitney test is based on an assumption of independent, identically-distributed errors. In the discussion following the examples, we examine the degree to which one must comply with these assumptions, and we comment on the proper interpretation of the results of hypothesis testing.

Example 4

Use the t-test to test the null hypothesis, at the 95% level, that the true mean chlorophyll a concentration in Lake B (μ_B) is identical to that for Lake C (μ_C) in Exhibit 5.

$$H_0: \mu_B - \mu_C = 0$$

$$H_1: \mu_B - \mu_C \neq 0$$

For this problem, Student's t is calculated from:

$$t = \frac{\bar{x}_B - \bar{x}_C}{\sqrt{\frac{s^2 (n_B + n_C)}{(n_B)(n_C)}}} \qquad (19)$$

where:

$\bar{x}_B$, $\bar{x}_C$ = mean chlorophyll a concentrations (µg/l) for lakes B and C (estimated from sample data)

n_B, n_C = the number of chlorophyll a observations for lakes B and C

s^2 = the pooled within-group variance (sample statistic).

To determine the pooled within-group variance, we must first calculate the sums of squares (ss) within each group.

$$ss_B = \Sigma x_B^2 - \frac{(\Sigma x_B)^2}{n_B} = (8.5)^2 + (4.2)^2 + \ldots + (10.2)^2 - \frac{(157.9)^2}{13} = 936.75$$

$$ss_C = \Sigma x_C^2 - \frac{(\Sigma x_C)^2}{n_C} = (65.6)^2 + (31.0)^2 + \ldots + (47.1)^2 - \frac{(410.8)^2}{13} = 3044.84$$

The pooled within-group variance is:

$$s^2 = \frac{ss_B + ss_C}{(n_B - 1) + (n_C - 1)}$$

$$s^2 = \frac{936.75 + 3044.84}{12 + 12}$$

$$s^2 = 165.9$$

Thus:

$$t = \frac{12.1 - 31.6}{\sqrt{165.9\ \frac{13 + 13}{(13)(13)}}}$$

$$t = -3.85$$

This value of t has $(n_B - 1) + (n_C - 1)$, or 24, degrees of freedom. Consulting a t-table (two-tailed), we find that for 24 degrees of freedom, this value of t is significant at the 99%+ level. This test supports rejection of the null hypothesis that the means are equal.

Example 5

Use the Mann-Whitney test to test the null hypothesis, at the 95% level, that the mean chlorophyll a concentration in Lake B is identical to that for Lake C in Exhibit 5.

$$H_0: \mu_B - \mu_C = 0$$

$$H_1: \mu_B - \mu_C \neq 0$$

The Mann-Whitney test is based on the W-statistic, which is the sum of the combined ranks occupied by the data points from one of the samples. The chlorophyll a observations are combined and ranked in Exhibit 7. To test the hypothesis, the ranks, R_i, associated with Lake B are summed.

$$W = \sum_{i=1}^{n_B} R_i \tag{20}$$

$$W \text{ (lake B)} = 110$$

At this point, the W-statistic may be compared to tabulated values to determine its significance. Alternatively, for moderate to large samples (n>10), W is approximately normal (if H_0 is true). This means that the W-statistic may be evaluated using a standard normal table and (Hollander and Wolfe, 1973):

$$W^* = \frac{W - E(W)}{[\text{var}(W)]^{\frac{1}{2}}} \tag{21}$$

or:

$$W^* = \frac{W - [n_A + (n_B + n_A + 1)/2]}{[n_B\ n_A\ (n_B + n_A + 1)/12]^{\frac{1}{2}}} \tag{22}$$

Exhibit 7. Chlorophyll a observation ranks for the Mann-Whitney test.

Combined Ordered Observations		Combined Ranks	
Lake B	Lake C	Lake B	Lake C
4.2		1	
4.7		2	
5.1		3	
6.5		4	
7.3		5	
8.2		6	
8.5		7	
	9.6		8
10.2		9	
12.1		10	
	14.2		11
15.0		12	
	19.6		13
20.4		14	
20.4		15	
	21.2		16
	23.0		17
	25.9		18
	30.0		19
	30.2		20
	31.0		21
35.2		22	
	42.1		23
	47.1		24
	51.3		25
	65.6		26

where:

W^* is N (0,1) when H_0 is true
W is calculated using Equation 20
E (W) is the expected value for the W-statistic
Var (W) is the variance for the W-statistic
n_A, n_B are the number of observations in samples A and B.

Since n_B, n_C > 10 for the problem posed, the significance of the W-statistic is determined using Equation 22.

$$W^* = \frac{110 - [13\ (13 + 13 + 1)/2]}{[(13)\ (13)\ (13 + 13 + 1)/12]^{\frac{1}{2}}}$$

$$W^* = -3.36$$

Consulting a standard normal distribution table (for a two-tailed test), it is found that this value of W^* is significant at the 99%+ level. The null hypothesis is therefore rejected. Note the similarity in test statistic values for the t-test and the W-test.

The assumptions inherent in the hypothesis tests, particularly in the t-tests, are cause for possible concern because they may be difficult to achieve. Fortunately, studies have been undertaken on the impact of violation of the assumptions. For example, Box et al. (1978) note that it is the act of "randomization" in experimental design, and not the use of a non-parametric technique, that makes a procedure insensitive to distribution assumptions. With randomization, Box and associates illustrate that both the t-test and the Wilcoxon test are relatively insensitive to the shape of the parent distribution, but they are both sensitive to serial correlation among the observations. In addition, Boneau (1962) found that the t-test is quite robust to violations of the assumptions of a normal parent distribution and of equal variances. Boneau concluded his study by noting that while the t-test should not be rejected because of concern over the aforementioned assumptions, neither should the Wilcoxon test be rejected because it is supposedly less powerful than the t-test. Both claims are sometimes false. The recommendation advanced here is proposed by Blalock (1972); apply both tests when in doubt about the assumptions. If the study is well-documented and the results of both a t-test and a Wilcoxon-Mann-Whitney test are reported, then the reader is provided with sufficient information for the analysis of the hypothesis.

In addition to concern over the assumptions, the user of an hypothesis test must be careful in the interpretation of the results. Specifically, an hypothesis test can be incorrect if we reject H_o when it is true (type I error), or if we accept H_o when it is false (type II error). The "significance level" (95% for the two examples) sets the probability of making a type I error. Since the significance level is known approximately, we know how often we are likely to reject H_o when it is true. However, type II error, evaluated by a test's "power," is dependent upon the true, but unknown, solution to the issue being tested. Therefore one cannot be certain of the likelihood of committing a type II error. There are power curve methods for estimating the probability of the type II error associated with true values for the issue being tested (see Wonnacott and Wonnacott, 1972). However, in the absence of these power determinations, the following recommendations are made. When the designated significance level is exceeded, the null hypothesis may be termed "rejected," and the significance level reported. Acceptance of H_o is another matter, however. When the alternative hypothesis covers a range of values (as in Examples 3 and 4), and the test statistic is not significant, then it is probably best to state that "H_o cannot be rejected." The alternative, "H_o is accepted" is too strong in the absence of power determinations. Additional testing would then be required, if a more definitive conclusion is needed.

Hypothesis testing is a confirmatory method in data analysis. The study of variable relationships may also occur in an exploratory mode as in certain graphical and statistical techniques for the analysis of bivariate data.

Among these techniques are correlation analysis, regression analysis, and bivariate plotting. Extension of the bivariate form of these techniques to multivariate data is straightforward but is not discussed here.

Correlation and regression analyses are frequently used in limnology for the examination of bivariate data. The correlation coefficient is a measure of the strength of a liner association, and it is an indicator of the predictive effectiveness of a regression equation. Regression analysis may be used to quantify the functional relationship (either linear or nonlinear) between two variables.

Most correlation and regression analyses are conducted with the aid of a calculator or digitial computer. It is unnecessary, therefore, to dwell on the mathematics of these techniques. The analyst of limnological data using one of these methods would be wise to devote some effort to understanding the assumptions inherent in regression and correlation analyses which may guide him/her in the interpretation of the results. For example, both regression and correlation are sensitive to trend outliers. As a result, robust methods have been proposed, in the form of rank-order correlation (Snedecor and Cochran, 1967) and robust regression (Reckhow and Chapra, 1980). Adherence to methodological assumptions is an important topic yet it is beyond the scope of this limited treatment. Therefore it is recommended that the analyst consult Reckhow and Chapra (1980), Kleinbaum and Kupper (1978), Wonnacott and Wonnacott (1972), Mosteller and Tukey (1977) or some other text that addresses the interpretation of correlation and regression relationships. Reckhow and Chapra (1980) provide an example illustrating how regression analyses can be quite misleading when the relationships are interpreted and applied, unless attention is paid to the assumptions.

In this brief presentation of data analysis, with an obvious emphasis toward concepts and robust methods, it seems appropriate to devote most of the bivariate relationships subsection to a discussion of bivariate plots. The analysis of bivariate relationships is quite common in limnological studies. For example, the trophic state index described in the next section is based upon three bivariate relationships among phosphorus concentration, chlorophyll a level, and Secchi disc depth. In the lake modeling field, modelers have debated the relationships between phosphorus concentration and mean depth, phosphorus concentration and areal water loading, and phosphorus concentration and hydraulic detention time. Often in these studies, correlation coefficients or regression equations are used in support of a bivariate relationship. The bivariate plot is also sometimes used, and it can be quite effective both in exploratory work to uncover relationships and in diagnostic work to study and check identified relationships. In fact it is recommended here that bivariate plotting be a standard feature of bivariate or multivariate data analysis. Reliance on statistics alone (e.g., on correlation coefficients only) can result in inaccurate analyses, as statistics can mask unusual data set characteristics that are quite evident when graphed (see Reckhow and Chapra, 1980).

Limnological data analysis has somewhat unusual features that might be studied using bivariate plots. Specifically, limnological data are often collected on a cross-section of lakes and then used to analyze relationships

in a single lake longitudinally, or over time. In the original cross-sectional analysis, each data point is not a single observation but rather a summary statistic (for location) representing several observations. So, there are two issues hidden in many bivariate limnological studies:

1. Is limnological behavior that is identified in a cross-sectional (multi-lake) analysis meaningful when applied to a single lake over time?

2. Is information lost when only summary statistics (for location) are used in (multi-lake) cross-sectional studies? If so, are there methods for recovering and examining this information while preserving the basic features of the cross-sectional study?

While we cannot provide a definitive answer to these questions (in part, because they are somewhat application-specific), an exploratory method related to the box plot yields some insight. It is based on a graphical analysis of the five order statistics (median, quartiles, and extreme values) employed in the box plots. As an example, the medians, quartiles, and extreme values are determined for the phosphorus data and for the chlorophyll a data presented in Exhibit 5. These statistics are then paired for each lake and plotted in Exhibits 8 and 9. In Exhibit 8, the five order statistics are connected for single lakes, while in Exhibit 9 the points are connected on the basis of matching statistics (medians with medians, etc.), across lakes. (Not all of the points in Exhibit 9 are connected by lines. A visual smoothing technique was employed to produce convex sections around the central tendency line. See Tukey, 1977, for simple mathematical methods for smoothing curves.)

There are a number of attributes of these plots worth exploring. First, the central tendency line (the line connecting the medians) in Exhibit 9 is equivalent to the standard trend line for cross-sectional regressions of chlorophyll a and phosphorus. The convex quartile and range lines surrounding the median provide an indication of variability to be expected within a single lake. Note that this is different from the scatter of data found in a cross-sectional regression, which is a function of variability among lakes. In Exhibit 8, the slope of the lines suggests the chlorophyll-phosphorus relationship within lakes. (Although, it must be remembered that the data points do not actually represent paired observations. Rather, the phosphorus and chlorophyll a data were ordered separately and then paired as order statistics, i.e., median with median.) A comparison of the slopes for single lake relationships (like in Exhibit 8) with the slope for multi-lake cross-sectional central trend is important. When the slopes are essentially equivalent, the multi-lake relationship is informative for single lake trend analysis. When the slopes are different, the multi-lake trend is misleading. In either case, the multi-lake variability (which represents cross-sectional differences, in part) and multi-lake prediction error are probably not too indicative of single lake variability. Thus predictive equations for bivariate relationships within lakes should probably be developed, when possible, from single lakes or highly homogeneous data for unbiased, minimum uncertainty predictions.

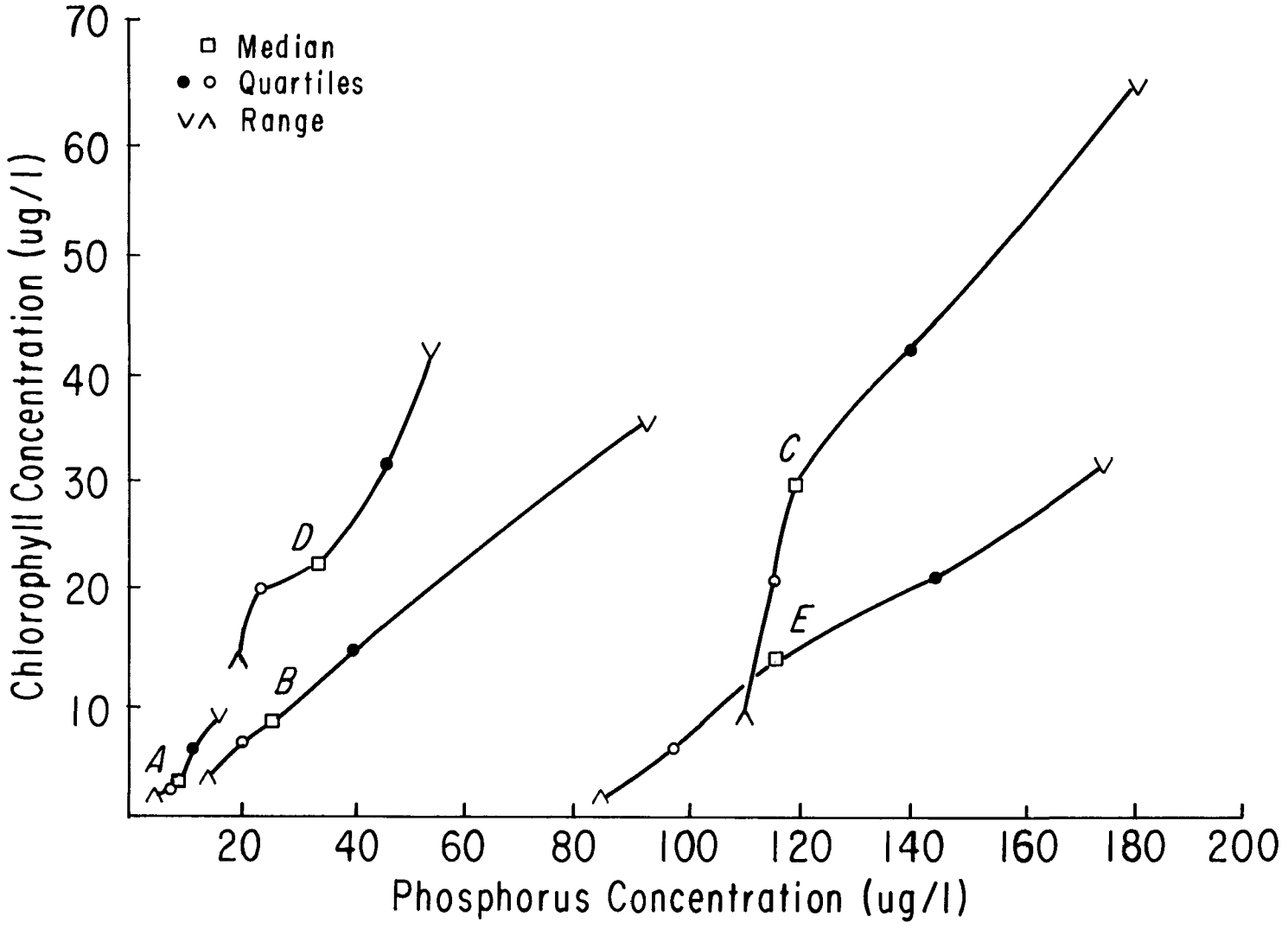

Exhibit 8. A bivariate plot: Single lake chlorophyll phosphorus relationships.

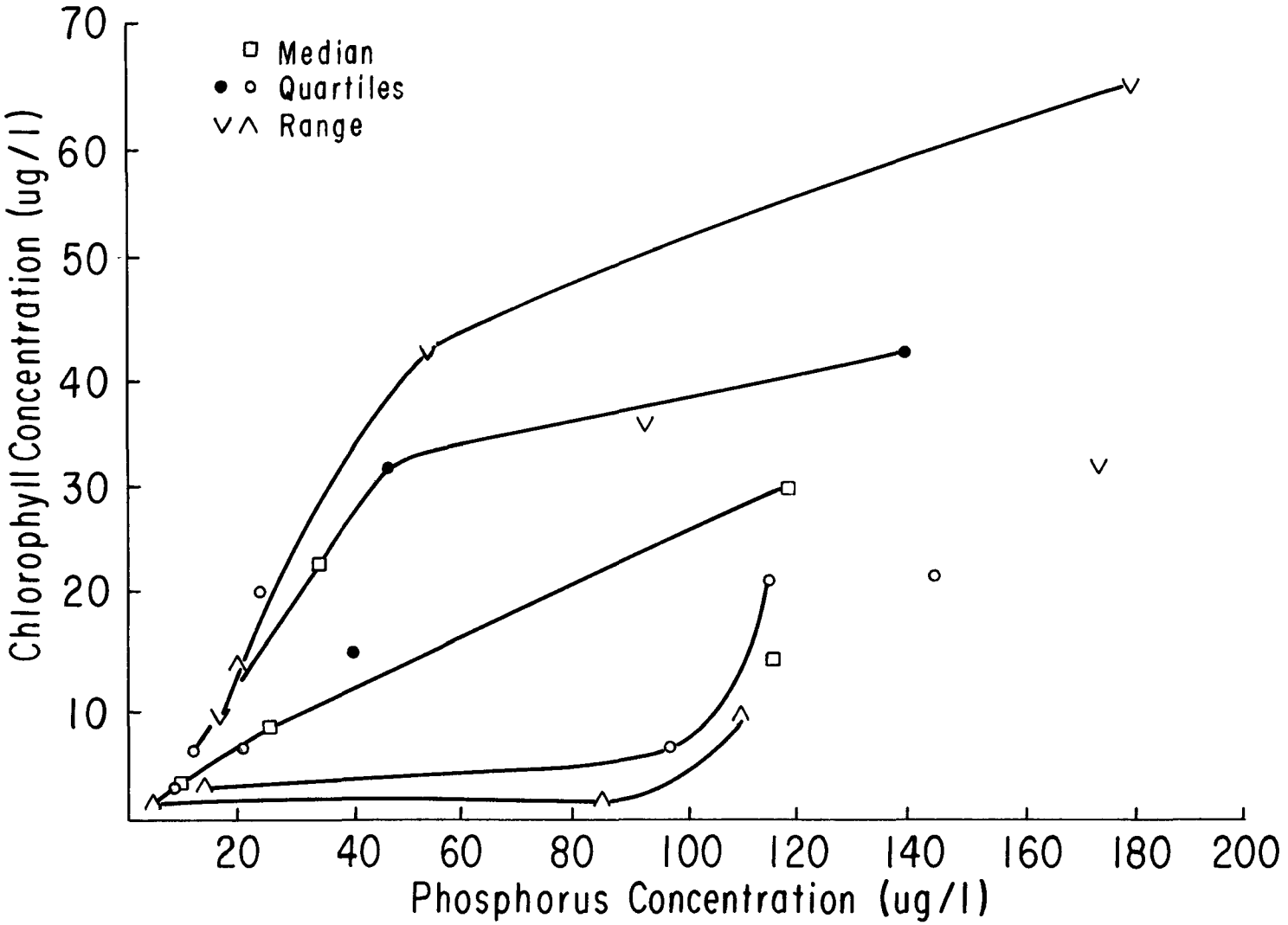

Exhibit 9. A bivariate plot: Cross-sectional chlorophyll phosphorus relationships.

Before ending this brief treatment of data analysis, some statistical references should be mentioned and briefly annotated. Reckhow and Chapra (1980) contain several chapters on data analysis and empirical modeling presented in a style and philosophy similar to the approach employed in this section. Tukey (1977) and Mosteller and Tukey (1977) are excellent references on exploratory data analysis, while Mosteller and Rourke (1973) and Hollander and Wolfe (1973) present nonparametric methods. Chatterjee and Price (1977) and Kleinbaum and Kupper (1978) are excellent in their treatment of applied regression analysis. Experimental design and other topics are covered in Box et al. (1978). Finally, Snedecor and Cochran (1967) and Wonnacott and Wonnacott (1972) are good, general references for several topics in statistics and data analysis.

In conclusion, three recommendations for data analysis should be apparent from the placement of the emphasis in this section.

1. Select summary statistics according to the vague concept criterion. That is, the statistic chosen to represent a data set should be the best choice because it represents the concept (e.g., location) best, and not because it is the natural choice in traditional statistical analyses (i.e., normal distribution theory).

2. When in doubt about the underlying distribution of a set of data, use robust statistics and methods.

3. The plotting of univariate, bivariate, and multivariate data is an essential step in statistical analysis.

4. Indices of Lake Water Quality

Considerable attention in the previous section was devoted to methods and statistics for summarizing data. Probably the most common of the summary statistics are the various measures of location, such as the mean, median, and mode. From an information perspective, we would call these location statistics univariate indices.

An index is a summary statistic. Since it is rarely a sufficient statistic, it contains less information than is available in the data set that it summarizes. A univariate index is a location statistic for a single variable, such as mean phosphorus concentration. A multivariate index is a single number chosen to summarize data on two or more variables. It is the multivariate index that is the focus of this section, although the univariate analogy is sometimes useful for discussion purposes.

Indices are used presumably because the convenience of summarizing information in a single number outweighs the disadvantage of information lost due to the act of summarization. It was pointed out in Section 1 that lakes provide for multiple uses which makes lake water quality a use-specific, or perhaps a problem-specific, attribute. A true water quality index, therefore, is multidimensional. The naturally subjective decisions as to which variables

should be part of a water quality index and what schemes should be used to combine the variables have been largely responsible for the dearth of widely used indices.

In this section, as in other sections of this discussion, we consider a specific lake quality problem: eutrophication. Now, the water quality index may be renamed a trophic state index (TSI). In addition, the index is reduced to essentially a single dimension associated with trophic state. Within this single dimension the index may still be multivariate, which means that variables are highly intercorrelated, representing the same basic concept (eutrophication).

A number of attempts have been made to establish a trophic state index as a function of commonly measured water quality variables. The EPA National Eutrophication Survey (1974) has compared the work of some investigators (Sakamoto, 1966; National Academy of Sciences, 1972; and Dobson et al., 1974) on chlorophyll a levels versus trophic state. This presented in Exhibit 10a. The EPA's own estimates of values of chlorophyll a, total phosphorus, and Secchi disc depth indicative of trophic states are presented in Exhibit 10b.

Exhibit 10a. Trophic state vs. chlorophyll a (from EPA-NES, 1974).

Trophic Condition	Chlorophyll a (µg/l)			
	Sakamoto	Academy	Dobson	EPA-NES
Oligotrophic	0.3-2.5	0-4	0-4.3	<7
Mesotrophic	1-15	4-10	4.3-8.8	7-12
Eutrophic	5-140	>10	>8.8	>12

Exhibit 10b. EPA-NES trophic state delineation (from EPA-NES, 1974)

Trophic State	Chlorophyll a (µg/l)	Total Phosphorus (µg/l)	Secchi Disc Depth (m)
Oligotrophic	<7	<10	>3.7
Mesotrophic	7-12	10-20	2.0-3.7
Eutrophic	>12	>20	<2.0

While there have been other attempts at single variable trophic state criteria (or indices), all are relatively similar in approach (see Exhibit 10). More importantly, they represent subjective judgment, and possibly limited geographic regions, so it is unlikely that universal agreement will rest on one approach. Therefore, the selection of a univariate trophic state criterion should be based primarily on personal acceptance and credibility.

More robust trophic state criteria or indices may be developed with a multivariate approach. Shannon and Brezonik (1972) constructed a trophic index for Florida lakes composed of the variables: primary production (PP, in mg, of carbon per cubic meter-hour), chlorophyll a (CA in mg/m^3), total organic nitrogen (TON, in mg/l as N), total phosphorus (TP, in mg/l as P), Secchi disc transparency (SC, in meters), specific conductance (COND, in μmho/cm), and a cation ratio (CR, a dimensionless ratio of (Na + K)/(Ca + Mg)). For lakes without appreciable organic color, the trophic state index (TSI) was estimated as:

$$\begin{aligned} TSI = {} & 0.936\ (1/SD) + 0.827\ (COND) + 0.907\ (TON) \\ & + 0.748\ (TP) + 0.938\ (PP) + 0.892\ (CA) \\ & + 0.579\ (1/CR) + 4.76 \end{aligned} \qquad (23)$$

A TSI of about 3 to 5 defines the transition zone between eutrophy and mesotrophy, and a TSI of 1.2 to 1.3 separates the mesotrophic and oligotrophic classes.

The index was developed using principal component analysis, and the TSI is the first principal component. This technique may be used to indentify "common elements" among variables, and the first principal component is a linear combination of the variables that best describes the most common element. When all of the variables in an analysis are thought to be good indicators of a concept called trophic state, then it is reasonable to assume that the most common element extracted from this set of variables (the first principal component) would be a good index of trophic state. In fact, this component is more "robust" than any one variable as an indicator of trophic state. This means that it is less likely, than a single variable index, to misclassify a lake based on an erroneous measurement. Incorrect data on one variable can lead to misclassification based on that variable, but it may not lead to misclassification if the classification criterion is based on other variables (correctly measured) as well.

Despite the fact that a principal component trophic state index has this desirable feature of robustness, the TSI proposed by Shannon and Brezonik cannot be recommended for use on north temperate lakes. The TSI was developed from a data base of Florida lakes only, and the significant climatic (and thus thermal) difference between that area and the north temperate region is likely to affect the index. Since this effect is unclear, we are unable to interpret the TSI in north temperate lakes. Equally important, most of the trophic variables are log-normally distributed, which means that the best estimate for the TSI should be made under a logarithmic transformation for these variables. Without this transformation (as in the case of Shannon's and Brezonik's TSI), the index may be biased and may appear misleadingly precise.

A trophic state index has been proposed by Carlson (1977) that may also be considered multivariate. Carlson's index may be estimated from summer values of Secchi disc depth (SD, in meters), summer total phosphorus concentration (TP, in mg/m^3) or summer chlorophyll a concentration (CA, in mg/m^3), or a weighted combination of all three. Carlson used regression

analysis to relate Secchi disc depth to total phosphorus concentration and to chlorophyll *a* concentration. He then reasoned that a doubling of biomass levels, or a halving of the Secchi disc depth, corresponds to a change in trophic state. Carlson assigned a TSI scale of 0-100 to the three trophic variables, such that a change of 10 units in TSI corresponds to a halving of the Secchi disc depth and a change in trophic state. The regression equations presented below then were used to relate the TSI to phosphorus and chlorophyll.

$$TSI = 60 - 14.41 \ln SD = XSD \quad (24)$$
$$TSI = 9.81 \ln CA + 30.6 = XCA \quad (25)$$
$$TSI = 14.42 \ln TP + 4.15 = XTP \quad (26)$$

Exhibit 11 contains the index values and variable relationships.

Exhibit 11. Carlson's trophic state index.

TSI	Secchi Disc (m)	Surface Phosphorus (mg/m^3)	Surface Chlorophyll (mg/m^3)
0	64	0.75	0.04
10	32	1.5	0.12
20	16	3	0.34
30	8	6	0.94
40	4	12	2.61
50	2	24	7.23
60	1	48	20
70	0.5	96	55.5
80	0.25	192	154
90	0.125	384	426
100	0.0625	768	1,180

Carlson's TSI may be estimated from any of the three variables, using Exhibit 11. Carlson felt that this was important as:

1. Secchi disc readings may be misleading as a trophic state indicator in colored lakes or highly turbid (non-algal) lakes.

2. Chlorophyll *a* may be the best indicator during the growing season.

3. Phosphorus may not be a good indicator in non-phosphorus limited lakes.

Thus different variables may be used depending upon the season, lake, and availability and quality of data. While Carlson suggests that the variable that the index is based on be selected on a pragmatic basis, he recommends that consideration be given to chlorophyll in the summer and to phosphorus in the fall, winter, and spring.

Recently, Porcella et al. (1980) have proposed a "Lake Evaluation Index" (LEI), based in part on Carlson's trophic state index, to be used to describe the effectiveness of lake restoration programs. The LEI, which Porcella et al. admit is still under development, is composed of 5-6 variables (all measured, preferably, between 1000 and 1400 standard time). They are:

1. Secchi Depth (SD). The LEI value (XSD), calculated in Equation 24, is based on the mean SD measured during the months of July and August. Color may be important (see below), so when present, it should be documented.

2. Total Phosphorus (TP). The LEI value (XTP) is calculated from the mean TP measured during July and August in the epilimnion. Equation 26 provides the index value.

3. Total Nitrogen (TN). At present nitrogen is not part of the LEI. However, a nitrogen index statistic has been determined from the mean TN measured during July and August in the epilimnion (in mg/m^3). This statistic is:

$$XTN = 14.427 \ln TN - 23.8 \quad (27)$$

4. Chlorophyll a (CA). The LEI value (XCA) is calculated from the mean CA measured during July and August in the epilimnion. Equation 25 provides the index value.

5. Dissolved Oxygen (DO). The LEI value (XDO) is equal to ten times net DO (in mg/l), which is calculated from July-August data.

$$\text{net DO} = \frac{\sum_{i=0}^{z_{max}} \left| (EDO - CDO)_i \right| \Delta V_i}{V} \quad (28)$$

where

z_{max} = maximum depth

i = index of depth contours

ΔV_i = volume of depth contour i

EDO = equilibrium DO, calculated from atmospheric pressure and temperature-depth profiles (kg/lake)

CDO = total lake DO (kg/lake).

Volume sections should be selected so that supersaturation and undersaturation do not cancel, if present, since they both are often indicative of quality deterioration. This can be accomplished by placing these "quantities" in different volume sections.

6. Macrophytes (MAC). The LEI value for macrophytes (XMAC) is defined as the percent area deemed "available" for macrophyte growth that is actually occupied by macrophytes. This available area is considered to be "the area encompassed by the lake margin and either the 10 meter line or the depth at which light becomes limiting to vascular plant distribution and growth (2 times SD) which ever is shallower" (Porcella et al., 1980).

The six "x-values" presented above convert the LEI variables to a 0-100 scale. These relationships are presented in Exhibit 11 for Carlson's variables and in Exhibit 12 for the other three variables. The actual LEI proposed by Porcella et al. is a composite variable, also on a 0-100 scale.

Exhibit 12. Rating scale for certain LEI variables (from Porcella et al., 1980).

Rating (X)	Total N (mg/m^3)	Net DO (mg/l)	Macrophytes % Available Lake Area Covered
0 (minimally impacted)	5.2	0.0	0
10	10.4	1.0	10
20	20.8	2.0	20
30	41.6	3.0	30
40	83.2	4.0	40
50	167.	5.0	50
60	333.	6.0	60
70	666.	7.0	70
80	1330.	8.0	80
90	2670.	9.0	90
100 (maximally impacted)	$\geq$5330.	>10.0	100

$$LEI = 0.25\ [0.5\ (XCA + XMAC) + XDO + XSD + XTP] \quad (29)$$

For both the LEI and the individual X-specified variables, an index value of less than 40-45 represents oligotrophy and an index value of greater than 50 represents eutrophy. Porcella et al. emphasize that the LEI is meant more as a measure of lake restoration effectiveness than as a trophic state index per se. Its usefulness, and linked to that--its acceptance, in either role remain to be seen.

Carlson's index is based around Secchi disc transparency. Recently, some investigators (Lorenzen, 1980; Megard et al., 1980; and Edmondson, 1980) have suggested that a bias may exist or result in Secchi disc-based cross-sectional trophic studies due to non-algal turbidity and color. Single lake longitudinal relationships (see Section 3) are recommended instead.

An alternative index system has been proposed by Walker (1979), who also recognized the problems inherent in Carlson's TSI due to non-algal light alternating factors. Walker's index is based around chlorophyll a, which is probably less influenced by non-biomass factors than is Secchi disc depth. The components of Walker's index are:

$$I_{CA} = 20.0 + 14.42 \ln CA \quad (30)$$

$$I_{TP} = -15.6 + 20.02 \ln TP \quad (31)$$

$$I_{SD} = 75.3 + 19.46 \ln (1/SD - \alpha) \quad (32)$$

where α is a term (m^{-1}) representing non-algal influence on transparency. Walker's index (I_W) is then simply:

$$I_W = (I_{CA} + I_{TP} + I_{SD})/3 \quad (33)$$

Walker used the index (I_W) in Equation 33 as an independent variable (along with lake mean depth, z) in a regression for hypolimnetic oxygen depletion rate (ΔHOD, in g/m^2-day). Re-arranging this regression equation to solve for the index value, Walker proposed the following trophic state index (now expressed as I_o):

$$I_o = 175 + 21.3 \ln (\Delta HOD) - 96.8 \ln z + 18.9 (\ln z)^2 \quad (34)$$

Walker derived and tested his indices on north temperate lakes. He found that the transition between oligotrophy and mesotrophy occurred between 25 and 30, the transition between mesotrophy and eutrophy occurred between 45 and 50, and the transition between eutrophy and hypereutrophy occurred between 65 and 70 (all values expressed in index units).

A comparison of the trophic state indices presented in this section suggests a great deal of similarity among them. This is not unexpected since the relationships among the trophic state variables (especially Secchi disc depth, chlorophyll level and phosphorus concentration) are well established and understood. The recent contributions (Carlson, Porcella et al., and Walker) present some new insight (e.g., inclusion of macrophytes and dissolved oxygen) and offer a 0-100 scale. This effectively eliminates the subjective labeling of oligotrophic, mesotrophic, and eutrophic states, and substitutes a scale easily interpretable by non-limnologists. This advantage may be

removed, however, by the tendency of researchers to mentally convert these index units back to one of the three standard tropic states for ease of interpretation. In summary, then the selection of a trophic state index from among those discussed in this section should probably be made on the basis of familiarity by the users, since no single index conveys appreciably more information than any of the others.

5. Acquisition of Nutrient Budget Data

A necessary step in lake quality management planning is an analysis of how present, and projected, watershed characteristics and activities affect water quality. Given the construction of most trophic state assessment schemes and the seasonal variability of nutrient sources, information on nutrient flux is most useful when it is acquired in yearly increments. When the issue of concern relates to present land use, the acquisition and examination of existing nutrient flux data or the sampling of nutrient sources on an annual basis is appropriate. When the latter course of action is chosen, the methods described in Section 2 and at the end of this section are useful.

Alternatively, water quality management planning for projected land uses necessitates an "indirect" assessment of the annual nutrient budget. Since measurements cannot be made for these nonexistent land uses, nutrient flux estimates must be determined from the literature reporting measurements taken at another location and/or time. Actually, the literature may be consulted on nutrient export coefficients for all nutrient budget assessments (present or projected). It must be noted, however, that use of non-application-specific data has an associated risk. That is, if the literature values are not representative of the application case, then bias is introduced into the analysis. This creates risk. When the analyst has a choice (e.g., when studying the impact of present land uses), the increased risk due to use of literature export coefficients must be evaluated against the increased cost of nutrient flux sampling. This is simply one of many situations in planning when expected outcomes need to be examined so that the trade-off between cost and risk may be analyzed. Clearly it is difficult to introduce much rigor or precision into this trade-off study. However, even some rough calculations of cost versus risk associated with alternative sources of nutrient budget data may greatly improve planning. See Reckhow and Chapra (1980), Chapter 1, for a discussion of this and other issues important in water quality modeling and planning studies.

The selection of appropriate nutrient export coefficients is a difficult task. Proper choice of export coefficients is a function of knowledge of the application lake watershed and knowledge of the watersheds of candidate export coefficients. It is through comparisons of these watersheds that the analyst arrives at the appropriate coefficients. Since a critical aspect of a watershed analysis/modeling exercise is the estimation of prediction error (see Section 6), the analyst should realize that poor choice of export values contributes to an increase in error. This contribution may be explicit or implicit in the analysis, depending upon whether or not the analyst is aware of all of the uncertainty introduced by his/her choice of phosphorus export

coefficients. Clearly, experience in the application of this modeling approach is a valuable attribute. Information on nutrient export coefficients is available in Reckhow et al. (1980) which contains both a presentation of candidate export coefficients and a description of the watershed characteristics for the candidate coefficients.

Direct assessment of a lake's nutrient budget or of the nutrient flux emanating from specific sources requires careful planning. Application of the sampling design relationships presented in Section 2, or of the concepts important in sampling design, can lead to efficient sampling programs based upon explicit trade-offs among different sampling schemes. In addition, an estimate of the uncertainty associated with carefully gathered data on nutrient flux is valuable information for use in the models and classification schemes presented in Sections 4 and 6.

Lake phosphorus budget sampling design is discussed in considerable detail by Reckhow (1978e). The remainder of this section contains a summary of some of the issues presented in that paper. Major sources of phosphous considered were tributaries, sewage treatment plants, urban runoff, precipitation, septic tanks - groundwater, and lake sediments. For each source, the sampling design was based on an estimation technique, or model, that converted the gathered data to an annual phosphorus flux estimate.

Concurrent with the design of a nutrient flux sampling program should be the consideration of nutrient flux estimation techniques. Flux may be estimated directly (as it is, generally, when the literature is consulted for phosphorus loading estimates), or it may be determined from separate assessments of nutrient concentration and volumetric water flow rate. The estimation of flux from concentration and flow data, in turn, may be accomplished in several ways (Reckhow, 1979e). Care must be observed in the flux estimation procedure because certain procedures favor certain sampling designs and because poor choice of estimation procedures can lead to bias and greater uncertainty in the nutrient loading estimate.

Phosphorus flux in lake tributaries has been studied extensively, and thus there is a substantial quantity of literature that may be used for the estimation of the expected magnitude and variability of that flux. The EPA National Eutrophication Survey is a good source of data, and many of the EPA-NES streams have been classified by land use (Omernik, 1977). In general, total phosphorus concentration (in streams) decreases with flow in streams impacted by a sizeable point source, and increases with flow in streams undisturbed by major point sources. On that basis, phosphorus flux is probably best estimated by multiplying average flow times the flow-weighted concentration or by a regression equation of flux on flow. Since those calculations of flux require information on flow, it is recommended that continuous flow measurements be made, or that a regression equation (of flow on precipitation and watershed characteristics) be used to provide flow data. Regression equations like that described are available from the U.S. Geological Survey. Sampling for concentration should be allocated among tributaries using stratified random sampling, and it should probably occur on 2-4 week intervals (with a random start, and allocated according to seasonal flow variations). More frequent sampling results in auto-correlation among

samples, and less frequent sampling may result in considerable error. Finally, some consideration should be given to sampling major storm events, as a large percentage of the phosphorus loading may occur during those times.

Much data also exist on wastewater treatment plants, and again the EPA-NES is a good source. Treatment plant data exhibit a distinct diurnal cycle, so composite sampling is preferrable. Phosphorus flux estimates may be made from flow-weighted concentration time flow (continuous flow data should be available). Existing EPA-NES data indicate that the average phosphorus concentration varies considerably from plant to plant, while the coefficient of variation of phosphorus concentration generally lies between .3 and .5. Sampling among plants should be based on stratified random design, while sampling over time should be based on random sampling to reach a desired or minimum precision.

Urban runoff sampling clearly must be geared to storm events. Insufficient data exist to guide sampling designs in most situations. Therefore, only some general recommendations can be made. Automatic sampling may be most effective, since human response to a storm may miss a portion of the "first flush." Composite sampling for concentration may be used to estimate flux, as average flow times average concentration. Grab sampling can be used to fit an exponentially-decaying concentration model (Marsalek, 1975), that may be used to estimate flux with continuous flow data.

Existing data on phosphorus in bulk precipitation (precipitation plus dry fallout) indicate considerable variability from year-to-year, site-to-site, and storm-to-storm. Bulk precipitation phosphorus results from industrial air pollution, bare agricultural fields, dirt roads, etc. In many lakes, precipitation is a relatively minor source of phosphorus. Thus, literature values for precipitation phosphorus (Reckhow et al., 1980) should probably be compared to the expected flux of phosphorus (to the lake of study) from other sources before a sampling program is undertaken for this source.

No satisfactory techniques have yet been developed to measure phosphorus flux to a lake from septic tanks and groundwater. The most common technique used is a soil retention coefficient, specific to a soil type. However, a constant soil retention does not consider the time-dependency of retention, the total volume of soil through which phosphorus in solution must pass, and the loading of phosphorus to the soil. Probably a better technique at this time is a system of "seepage meters" in the shallow lake sediments and wells immediately onshore (Lee, 1977). The seepage meters are used to measure groundwater flow (assumed to decrease exponentially with distance from shore), and the wells are used to measure phosphorus concentration. Unfortunately, insufficient data exist to design this program, but the concepts of stratified random sampling (magnitude, variability, and cost) suggest that sampling units should be most dense in areas with the greatest density of septic tanks and in areas with soils of lowest retention coefficients.

Finally, the lake sediments are another source of phosphorus that is not well-defined. As a rule, the sediments are considered to be a significant source only under anerobic conditions. However, studies indicate (Snow and DiGiano, 1976) that aerobic sediments often release phosphorus also. Esti-

mation techniques, such as a constant daily release of phosphorus, or release proportional to the concentration gradient between the water column and the interstitial water, have been proposed (Reckhow, 1978e). Experimental procedures have been developed for both the laboratory and the field (Snow and DiGiano, 1976). It is suggested that "typical" release rates, presented in Reckhow (1978e) and Snow and DiGiano (1976), be compared to expected phosphorus flux from other sources, before a sampling program is undertaken for the lake sediments.

As an example of lake phosphorus budget sampling design, the following analysis was conducted to guide the sampling of phosphorus flux to Lake Winnipesaukee in New Hampshire (Reckhow and Rice, 1975; Reckhow, 1978e). This analysis emphasized the concepts of stratified random sampling (base a sample design on flux magnitude, variability, and sample cost); it did not consist of the explicit trade-offs and computations that might be possible with the material presented above. Nonetheless, it does show, in a general sense, how sampling design may develop.

Exhibit 13 presents the mean, standard deviation, and precision of existing estimates for phosphorus flux from the tributaries and the wastewater treatment plants. Informed judgment yielded the magnitude and range estimates for the other three phosphorus sources; data were deemed insufficient to specify these terms more precisely. This table, then, provided the basis for general sampling design recommendations, summarized in the following statements.

Exhibit 13. Initial uncertainty estimates for Winnipesaukee phosphorus loading.

	Prior Estimates			
Term	Magnitude	Coefficient of Variation	Standard Error of the Mean (%)	Estimated Range
1. Tributary Flux	16,000 lb P/yr	.65	±20	
2. Septic Tanks				4,000-30,000 lb P/yr
3. Sewage Treatment Plants	22,000 lb P/yr	.30	±10	
4. Precipitation				4,000-7,000 lb P/yr
5. Sediment Release				700-7,000 lb P/yr

1. Existing estimates of the phosphorus flux from tributaries and sewage treatment plants may be sufficient (i.e., no additional sampling necessary), if they were obtained with an unbiased sampling design, and if significant changes (land use, etc.) have not occurred.

2. Considerable sampling effort should be devoted to estimating the mean and variance in phosphorus flux from septic tanks.

3. The other sources of phosphorus (sediments and precipitation) should be investigated through the literature, but they may not require sampling.

4. If tributary sampling is undertaken:

   a) spatial coverage should be based on stratified random sampling design (which may result in no sampling in the smallest streams that are not culturally impacted).

   b) temporal coverage should consist of a sampling interval of 2-3 weeks, with sampling being more frequent during high runoff months and less frequent during months of low runoff.

In conclusion, a good sampling design requires the following information:

1. Prior knowledge of the factors that affect the characteristic(s) to be sampled (e.g., sources of phosphorus for a phosphorus budget).

2. Some knowledge of the magnitude and variability of the characteristic(s) to be sampled.

3. Pre-specified needs for the collected data. For example, phosphorus flux data may be used for a year-of-sampling estimate or for future predictions. Different designs and estimation techniques may be appropriate for each of these applications.

4. A knowledge of costs associated with sampling.

5. A model, or models, for estimation (when appropriate) that is compatible with the chosen sampling design.

6. Lake Trophic Quality Modeling

The prediction of the impact of watershed characteristics and activities on water quality is a necessary task in successful lake water quality management planning. Prediction implies the use of a conceptual, and most likely, mathematical, model to express variable relationships and make projections. To this end, many mathematical models have been developed and proposed for lake trophic quality management. Initially, most of these models were presented in a deterministic mode. However, as modelers acquired more information on the functioning of a lake and watershed system, and as

engineers and planners inquired about the reliability of the models, considerations of uncertainty began to appear. Modelers who examined uncertainty in their models, and planners who demanded an estimate of the uncertainty in the techniques that they used, realized that they must have a measure of the reliability of their methods. Without this, there was no way to assess the value of the information provided by a model. Under those conditions, inefficient or incorrect decisions were more apt to be made because the model results were given too much or too little weight.

Despite the fact that there are many water quality models in existence and more being developed, this does not necessarily represent a significant duplication of effort. Models are needed for a range of problems, and thus they are developed to address a variety of issues at different levels of mathematical complexity and for different degrees of spatial and temporal resolution. Thus, for a model user, the choice of model to be applied will depend upon:

1. the issue of concern,

2. the level of spatial and temporal aggregation appropriate to the issue,

3. the familiarity of the users to a particular model, or the mathematical sophistication of the user,

4. the cost and time required for acquisition of data necessary to run the model, and

5. the cost of model acquisition and model runs.

In the field of lake trophic quality modeling, ecosystem models (Thomann et al., 1975; Scavia and Robertson, 1979) have been developed to address the problem of eutrophication in a multi-dimensional manner, often with a fairly high degree of spatial and temporal resolution. In order to make these models more useful in the planning process, modelers have begun to quantify the error terms for ecosystem models (Scavia, 1980). As this occurs, lake ecosystem models will become more useful for the evaluation of lake management strategies.

At the other end of the lake model complexity spectrum, black box nutrient models have been proposed for the assessment of certain lake quality issues where considerable spatial and temporal aggregation is permissible. These models are attractive to many planners and engineers because they are often more compatible with the position of the planner/engineer on the model selection criteria mentioned above (particularly with regard to mathematical background and financial support). Since it has been shown that uncertainty analysis is relatively easily applied to the black box model, modeling with error analysis is now being undertaken by a group of model users who might otherwise work strictly with deterministic methods.

This is not to say that all lake model users addressing management concerns should be applying black box models. On the contrary, the first and second model selection criteria identified above clearly state that the chosen model should be appropriate to the issue of concern. Certainly there are many issues of importance in lake quality that are not addressed well with the black box model. Yet, at the same time, there are issues, and potential model users, who need simple, aggregated models, because of model selection criteria 3, 4, and 5. Some of these users may demand an estimate of the model uncertainty. It is more likely, however, that many of these users may not have thought a great deal about uncertainty. A procedure that allows these individuals to calculate a numerical value for an estimate of prediction uncertainty can be a powerful tool for convincing engineers, planners, and decision makers of the value of uncertainty. Therefore the emphasis in this section is on a discussion of black-box lake models and associated error analyses.

Empirically-based input-output lake models for phosphorus were first proposed in the early 1960s (see Reckhow, 1979a). However, management and planning applications of these methods were most stimulated by Vollenweider's thorough analysis (1968) in which he suggested nutrient loading criteria for lakes as a function of mean depth. In the past twelve years, several variations (Vollenweider, 1975, 1976; Dillon and Rigler, 1975; Chapra, 1975; Larsen and Mercier, 1976; Jones and Bachmann, 1976; Reckhow, 1977, 1979b; Walker, 1977; Rast and Lee, 1978) of this basic theme have been proposed. These variations have the common features that: (1) they were developed from a cross-sectional analysis of lake data on annual phosphorus loading, phosphorus concentration, and selected hydrologic and geomorphologic variables; (2) empirical "curve-fitting" (objective or subjective) techniques were used on the cross-sectional data base to relate phosphorus concentration (sometimes equated with trophic state) to the other variables; and (3) for a single lake, the methods developed all describe a constant proportional relationship (expressed in terms of the hydrologic/geomorphologic variable(s)) between annual phosphorus loading and "average" lake phosphorus concentration. These methods are sometimes expressed graphically (e.g., phosphorus loading criteria) and sometimes expressed in equation form. Essentially the same information is conveyed in either case, so the choice among presentation modes is largely dictated by the needs of a particular application.

Probably the major difference among the input-output models (and graphical procedures) is the variation among the cross-sectional data bases used to estimate the model parameter(s) (or to locate the trophic state transition lines). As Reckhow (1979a) notes, some of the models were empirically fitted on a homogeneous data base and are uncorroborated for use on lakes with characteristics different from those of the model development data set. This could result in prediction bias in the uncorroborated cases. On the other hand, models developed from a homogeneous data base often have smaller standard errors than do heterogeneously-based models. As a general rule, a preferred model is one developed using a homogeneous data base from the subpopulation of lakes containing the application lake(s). In that situation, some exogeneous variables, important in a heterogeneous data base, are effectively "controlled for" by reducing lake type variability within the model development data set.

In mathematical terms, the input-output phosphorus lake model may be expressed in three basic forms:

$$P = \frac{L_\tau}{z}(1-R) \tag{35}$$

$$P = \frac{L}{v_s + q_s} \tag{36}$$

$$P = \frac{L}{\sigma_z + q_s} \tag{37}$$

where, on an annual basis,

P = lake phosphorus concentration (mg/l)

L = annual areal phosphorus loading (g/m^2-yr)

z = lake mean depth (m)

τ = hydraulic detention time (yr)

R = lake phosphorus retention coefficient (dimensionless)

q_s = areal water loading (m/yr)

v_s = apparent settling velocity (m/yr)

σ = sedimentation coefficient (yr^{-1})

$\frac{L_\tau}{z}$ = average influent phosphorus concentration (mg/l).

The model parameters are R, v_s, and σ, respectively. Traditionally, v_s and σ have been estimated by constants, while R has been fitted as a function of q_s or τ. Comparisons (Reckhow, 1979a, and Reckhow and Chapra, 1980) among the fitted models have been made to indicate lake types for which the models are in relative agreement or disagreement.

An example of the graphical form of the intput-output models in Vollenweider's phosphorus loading criterion relating L and q_s. A version of the loading criterion is presented in Exhibit 14, with the solid lines distinguishing oligotrophic, mesotrophic, and eutrophic states. The dashed and dotted lines reflect the model estimation error associated with the prediction of trophic state (set equivalent to the phosphorus trophic state

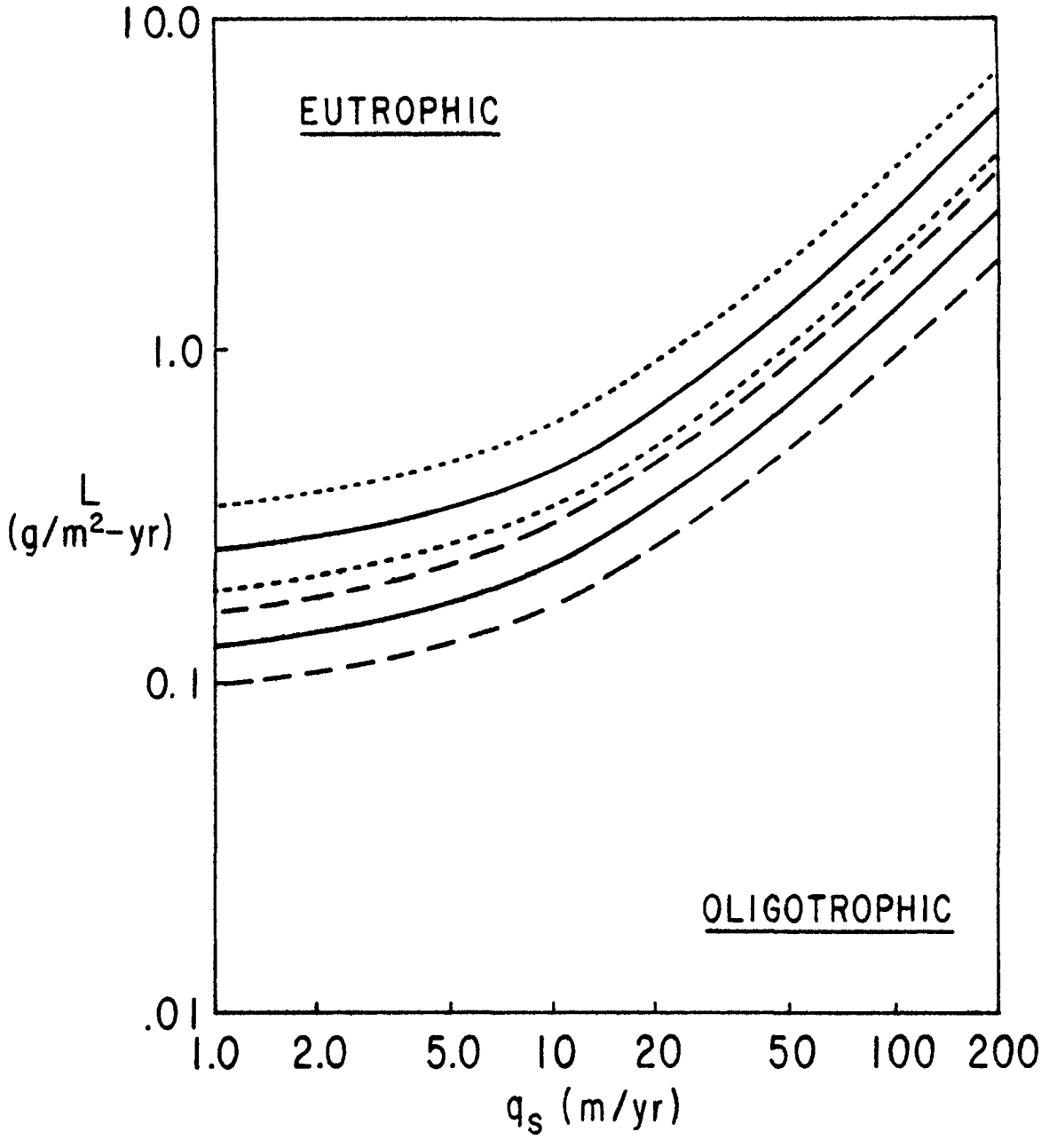

Exhibit 14. Vollenweider's phosphorus loading criterion with model estimation error.

criterion in Exhibit 10b) from L and q_s. This is only a portion of prediction uncertainty for most applications, and Reckhow (1979d) proposes a graphical method for estimating the magnitude of the additional uncertainty. An alternative methodology for complete model prediction uncertainty estimation is presented below. First, a short discussion of uncertainty is in order, however.

There is always uncertainty in the prediction of a model. Quantification of this uncertainty can be a useful exercise, because the level of uncertainty is inversely related to the value of the information contained in the prediction. Uncertainty in modeling arises from three primary sources: the input data for the model, the model parameters, and the model itself. One approach that may be used to estimate prediction uncertainty is first order error analysis (Cornell, 1972). Under this method, the error in a characteristic (variable or parameter) is defined by its first nonzero moment (the variance). Errors are propagated through the model using the first order terms in the Taylor series, and the variances are then combined to yield the total prediction uncertainty.

An alternative approach to model prediction error analysis is Monte Carlo simulation. Under this technique, probability density functions are assigned to each characteristic (variable or parameter), reflecting the uncertainty in that characteristic. Then, under the Monte Carlo procedure, values are randomly selected from the distribution for each term. These values are inserted into the model, and a prediction is calculated. After this is repeated a large number of times, a distribution of predicted values results, which reflects the combined uncertainties.

The quantification of uncertainty associated with the application of lake models is a relatively recent development. Apparently the first work on this topic was undertaken by Reckhow (1977) and by Walker (1977). In the past few years, Reckhow (1979abcd), Chapra and Reckhow (1979), and Reckhow and Chapra (1979) have expanded upon the use of first order error analysis with input-output lake models. Much of this is summarized in Reckhow and Chapra (1980). In addition, O'Hayre and Dowd (1978), Duckstein and Bogardi (1978), Reckhow et al. (1980a), and Montgomery et al. (1980) have employed Monte Carlo simulation to quantify lake modeling errors. Of these four, the first two have proposed a Bayesian approach. Recently, Scavia (1980) described work on quantifying lake ecosystem model prediction error using Monte Carlo simulation and the Extended Kalman Filter (a dynamic counterpart for first order analysis). This is among the first attempts at error analysis for a relatively complex lake model.

Five years ago, Dillon and Rigler (1975) proposed a step-by-step procedure for the estimation of lake phosphorus concentration using a simple input-output model. When employed in prediction or lake quality management planning, the methodology included steps for the selection of annual phosphorus export coefficients (see Uttormark et al., 1974) associated with each land use. This procedure has proven to be quite popular as a relatively comprehensive guide to the use of nutrient export coefficients and an input-output lake model.

One important feature missing from the Dillion-Rigler methodology is a step for the estimation of prediction uncertainty. Therefore, a procedure has recently been proposed (Reckhow and Simpson, 1980) that includes a step describing the estimation and combination of errors for the calculation of a nonparametric prediction interval. This procedure employs a phosphorus lake model of the form presented in Equation 36. Using nonlinear least squares, the model parameter, v_s, was estimated (Reckhow, 1979d) as:

$$v_s = 11.6 + 0.2\ q_s \tag{38}$$

resulting in the empirical phosphorus lake model:

$$P = \frac{L}{11.6 + 1.2\ q_s} \tag{39}$$

The Reckhow-Simpson procedure is described in step-by-step detail in the original reference and elsewhere (Reckhow et al., 1980b; Reckhow and Chapra, 1980), so an overview of the phosphorus loading estimation methods are presented below followed by a detailed explanation of the error calculation steps. For phosphorus loading determination, it is recommended that high, most likely, and low export coefficients be selected for the phosphorus source categories. This allows the calculation of high, most likely, and low total loading estimates. The high and low loading estimates represent the additional phosphorus loading error that must be added to the model error for the calculation of total prediction uncertainty. It is important that the high and low loadings primarily represent uncertainty due to (1) projection uncertainty associated with anticipated land use and population changes during the planning period, and (2) extrapolation uncertainty associated with the use of phosphorus export data measured at another point in space and/or time. This requirement exists because to a great extent, the error in the phosphorus loading estimates is already contained in the model error. Additional loading error for an application lake must be included only when the loading is estimated (using the procedure herein) in a different (and less precise) manner than it was estimated for the model development data set. The references mentioned above offer additional guidance in the choice of phosphorus export coefficients, and Exhibit 15 presents some typical values.

The selection of appropriate phosphorus export coefficients is a difficult task. It is largely contingent upon the analyst matching the application lake watershed with candidate export coefficient watersheds according to characteristics that determine phosphorus export from the land. A close match should insure that the selected export coefficients are reasonably representative of conditions in the application lake watershed. Since a critical aspect of this modeling exercise is the estimation of prediction errors, the analyst should realize that poor choice of export values contributes to an increase in error. This contribution may be explicit or implicit in the analysis, depending upon whether or not the analyst is aware of all of the uncertainty introduced by his/her choice of phosphorus export coefficients. Clearly, experience in the application of this modeling approach is a valuable attribute. Following selection of phosphorus export coefficients and calculation of the total phosphorus loadings, the three total phosphorus loading estimates are then separately inserted into the model

Exhibit 15. Phosphorus export coefficients (units are Kg/$10^6 m^2$-yr, except septic tank as indicated; values are adopted from Uttormark et al., 1974, and Reckhow et al., 1980).

	Agriculture	Forest	Precipitation	Urban	Input to Septic Tank (Kg/capita-yr)
High	300	45	60	500	1.8
Mid	40-170	15-30	20-50	80-300	0.4-0.9*
Low	10	2	15	50	0.3

* The value selected will depend, in part, upon whether or not phosphate detergents are permitted.

(Equation 39), and "high," "most likely" and "low" ($P_{(high)}$, $P_{(ml)}$, and $P_{(low)}$ respectively) lake phosphorus concentrations are calculated.

In order to estimate the uncertainty associated with a prediction calculated using the phosphorus model, estimates are needed for the error, or uncertainty, in all terms in the model, and in the model itself. However, it has been shown by Reckhow (1979b) that for most applications of this model, the error in the parameter v_s is small. Further, error in q_s is primarily a function of flow measurement error and hydrologic variability, which also affect L. Since L and q_s are in the numerator and denominator, respectively, in the model, the errors affecting both tend to cancel when they are combined to yield the resultant error in P. In addition, hydrologic variability is unimportant in lakes with low flushing rates. Therefore, it is assumed here that the prediction error is a function only of model error and of aspects of phosphorus loading uncertainty that are identified in Reckhow and Simpson (1980). If the application lake flushes rapidly and is subject to great variations in year-to-year precipitation, then the modeler is urged to include hydrologic variation in the error analysis using the error propagation equation (Reckhow et al., 1980, outlined the appropriate procedure).

The model error is represented by s_{mlog} in the equations below and is expressed in logarithmic units of phosphorus concentration error. The loading error, s_L, on the other hand, is expressed in untransformed units of phosphorus loading error. Therefore, to combine these two values for an estimate of total prediction uncertainty, some calculations are necessary.

The procedure presented below is based on first order error analysis (Benjamin and Cornell, 1970). In this particular application, three assumptions are of some importance:

1. Model error, expressed in log-transformed concentration units, is appropriately combined with variable error terms after the transformation is removed.

2. The "range" ("high" minus "low"), for phosphorus loading error, is approximately two times the standard deviation. This is based loosely on the characteristics of the Chebyshev inequality identified below, where about 90% of the distribution is contained within ±2 standard deviations of the mean.

3. The individual error components are adequately described by their variances (standard deviations).

In order to relax a previously imposed (Reckhow, 1979b) yet tenuous normality assumption, the confidence intervals constructed below are based on a modification of the Chebyshev inequality (Benjamin and Cornell, 1970). Therefore, it is no longer required that the total error term be normally distributed. Instead its distribution must only be unimodal and have "high order contact" with the abscissa in the distribution tails. These are achievable assumptions under almost all conditions, and it is recommended that this type of nonparametric approach be adopted until the distributions have been adequately studied and characterized.

Step A: Calculation of log $P_{(ml)}$

Take the logarithm of the most likely phosphorus concentration, $P_{(ml)}$.

Step B: Estimation of s_m+ ("positive" model error)

The model error, (s_{mlog}), was determined to be 0.128. Add s_{mlog} to log $P_{(ml)}$ and take the antilog of this value. Now calculate the difference between this antilog value and $P_{(ml)}$). Label this difference s_m+; it represents the "positive" model error.

$$s_m+ = \text{antilog} [\log P_{(ml)} + s_{mlog}] - P_{(ml)} \quad (40)$$

Step C: Estimation of s_m- ("negative" model error)

Subtract s_{mlog} from log $P_{(ml)}$ and take the antilog of this value. Now calculate the difference between this antilog and $P_{(ml)}$, and label this difference s_m-.

$$s_m- = \text{antilog} [\log P_{(ml)} - s_{mlog}] - P_{(ml)} \quad (41)$$

Step D: Estimation of s_L+ ("positive" loading error)

Now, one must convert the loading error estimate into units compatible with the model error. Use the $P_{(high)}$ concentration estimated earlier and

calculate the difference between $P_{(high)}$ and $P_{(ml)}$; then divide this difference by 2. Label this value s_L^+; it represents the "positive" loading error contribution.

$$s_L^+ = \frac{P_{(high)} - P_{(ml)}}{2} \tag{42}$$

Step E: Estimation of s_L^- ("negative" loading error)

Repeat Step D substituting the low concentration value $P_{(low)}$ for $P_{(high)}$. Label the resultant value s_L^-; it represents the "negative" loading error contribution.

$$s_L^- = \frac{P_{(ml)} - P_{(low)}}{2} \tag{43}$$

Step F: Estimation of s_T^+ (total "positive" uncertainty)

Total positive prediction uncertainty is calculated using the equation:

$$s_T^+ = \sqrt{(s_m^+)^2 + (s_L^+)^2} \tag{44}$$

Step G: Estimation of s_T^- (total "negative" uncertainty)

Total negative prediction uncertainty is calculated using the equation:

$$s_T^- = \sqrt{(s_m^-)^2 + (s_L^-)^2} \tag{45}$$

Step H: Calculation of confidence limits.

The prediction uncertainty may be expressed in terms of "confidence limits" which represent the prediction plus or minus the prediction uncertainty. Confidence limits have a definite meaning in classical statistical inference; they define a region in which the true value will lie a pre-specified percentage of the time.

Using the modification of the Chebyshev inequality (Benjamin and Cornell, 1970), the confidence limits may be written as:

$$\text{Prob}\left[P_{(ml)} - hs_T^- \right) \leq P \leq \left(P_{(ml)} + hs_T^+\right) \geq 1 - \frac{1}{2.25h^2} \tag{46}$$

Equation 46 states that the probability that the true phosphorus concentration lies within certain bounds, defined by a multiple, h, of the prediction error, is greater than or equal to $1-1/2.25h^2$. (This relationship loses its significance as h drops much below one.) Substituting values for h into Equation 46 reveals that a value of one for h corresponds to a probability of about 55% (.556 to be exact), and a value of two for h corresponds to a probability of about 90% (.889 to be exact). Thus the 55% confidence limits are:

$$\text{Prob}\ [(P_{(ml)} - s_T^-) \leq P \leq (P_{(ml)} + s_T^+] \geq .55 \qquad (47)$$

Once specific values for the prediction error have been inserted into the confidence limits expression, its interpretation changes somewhat. It is: "about 55% of the time (that confidence limits are estimated), one can expect that the actual phosphorus concentration will lie within bounds defined by the prediction plus or minus the prediction uncertainty." This same interpretation format applies when the confidence limits are widened to the 90% level (h=2), and specific data are inserted:

$$\text{Prob}\ [(P_{(ml)} - 2s_T^-) \leq P \leq (P_{(ml)} + 2s_T^+)] \geq .90 \qquad (48)$$

7. Concluding Comments

Mathematical models and statistical methods can be quite helpful for the analysis of quantitative problems. When used incorrectly, however, these techniques can yield misleading results that ironically have high credibility due to their mathematical or statistical basis. Therefore, it is important that the analyst understand the inherent assumptions, the limitations, and the proper use of the methods presented herein. To underscore some of the issues concerning the use of the models and statistics, some concluding thoughts are offered below:

1. There are certain procedures to be followed in scientific studies, and these procedures are collectively called the scientific method. Analysts engaging in scientific endeavors should be cognizant of proper definition of vague terms, specification of assumptions inherent in their work, considerations of uncertainty and risk, causality, and testing or corroboration of models.

2. The acquisition of data is frequently a problem of statistical sampling design. Often the design choice reflects a trade-off between the cost of analysis and the resultant uncertainty associated with the acquired data. There are both concepts and mathematical relationships that can be helpful in designing these programs.

3. Data analysis should be undertaken with consideration of the "vague concept" of interest. Graphical analysis of data is often helpful.

4. Since phosphorus loading/lake response modeling is probably a principal concern to users of this document, several comments are presented on this topic:

a. Reckhow (1979d) and Reckhow and Simpson (1980) identify the major application limitations for the modeling/uncertainty analysis procedure presented in the last section. In fundamental terms, the limitations are generally associated with the fact that the model development data set for any particular model represents a subpopulation of lakes. Application lakes that differ substantially from the model development subpopulation may not be modeled well (i.e., results may be biased). Any limnologic characteristic that is a causal determinant of lake phosphorus concentration is a candidate as a limiting, or constraint, variable. These include constraints on the model variables (e.g., all model development data set lakes have $P < .135$ mg/l), constraints on hydrology (e.g., there are no closed lakes in the model development data set), or constraints on climate (e.g., the model development data set contains only north temperate lakes).

b. The methodology described in the previous section can be used to quantify the relationship between watershed land use and lake phosphorus concentration. Yet phosphorus by itself is not an objectionable water quality characteristic. The real quality variable of concern (i.e., the characteristic(s) that lend(s) value or human benefit to the water body, abbreviated "qvc") may be algal biomass, water clarity, dissolved oxygen levels, or fish populations. Therefore the modeling methodology and the error analysis do not include all of the calculations necessary to link control variables (land use) with the qvc. This means that the relevant prediction error (on the qvc) is underestimated by the phosphorus model prediction error, and planning and management risks are inadequately specified. More useful methodologies are needed that quantitatively link control variables with the qvc for a particular application.[1]

c. The error analysis procedure suggested by Reckhow and Simpson should provide a reasonable estimate of prediction uncertainty. However, there are still problems in interpretation and application. For instance, the model error component was estimated from a least squares analysis on a multi-lake (cross-sectional) data set. This error is then applied to a single lake in a longitudinal sense. Thus, much of the model error term actually results from multi-lake variability, whereas when the model is applied to a single lake, the model error term should consist primarily of lack-of-fit bias and single lake variability. On the basis of present knowledge, it

[1] Certain complex models (see Scavia and Robertson, 1979) are comprehensive in system coverage from control variables to qvc. However, these models possess other shortcomings (large error terms and inadequate testing or corroboration) that affect their utility in lake quality management planning.

is not clear how a multi-lake-derived error relates to a single lake analysis.

d. A second issue associated with the error analysis concerns the subjective determinations of phosphorus loading and hence, loading estimation error. Statisticians and modelers generally prefer objective measures of uncertainty, such as calculated variability in a set of data. However, both limited available data and the obviously unmeasurable nature of future impacts favor (or necessitate) subjective estimates. Given this subjectivity, and the inexperience of most planners and analysts with phsophorus loading estimation, there may be uncertainty in the uncertainty estimates. This is exacerbated by the potential for loading error "double counting" (see Reckhow, 1979d), although the Reckhow/Simpson procedure is designed to reduce error double counting. It is likely that as analysts gain experience in loading and error estimation, this problem will be of less importance.

e. At this time a comment on model selection is in order, given the number of models developed in recent years. It is probably presumptuous of a modeler to label his/her model as "best" without stating some relevant qualifications or criteria. A "best" model is generally best according to some error criterion (like least squares) and for some subpopulation of the cases modeled. The planner/analyst should select a model that has been documented as best for conditions identical or similar to those of concern. Reckhow and Chapra (1980) discuss several characteristics and criteria that should be included in the model developer's documentation of his/her water quality management planning model. The prospective model user would be wise to request and examine this documentation before selecting the application-specific best model.

5. Ultimately the analyses conducted under the guidance of this document will be used to aid lake quality management planning. Therefore, given this planning objective, two final thoughts are offered for the analyst to consider:

a. Water quality management planning and modeling incur a cost that is presumably justified in terms of the value of the information provided. The actual achievement of a water quality level often requires management and pollutant abatement costs but also carries with it various benefits. The analyst must be cognizant of the fundamental economic nature of environmental management, planning, and decision making. The acquisition of additional data or the conduct of additional modeling and planning studies should be justified in terms of information return for improved decision making.

b. The planner or analyst conducting a lake data analysis or modeling study has as his/her primary goal the effective com-

munication of the work carried out. This does not simply mean documentation of the calculations and presentation of the statistics or the prediction and prediction uncertainty. Rather, effective communication requires consideration of the knowledge and concerns of the likely audience. The analyst must then describe his/her study so that the audience can comprehend the results, can understand the study's limitations, and can act (if necessary) in an informed manner. As a rule, this means that the analyst should completely describe procedural limitations and assumptions made in conducting the study. Beyond that, the analyst should explain how the limitations and assumptions affect the interpretation of the results for planning. A comprehensive discussion of the application of the statistical analysis or the modeling methodology that meets the needs of the intended audience facilitates good water quality management planning.

References

Ackoff, R. L. 1962. Scientific Method: Optimizing Applied Research Decisions. J. Wiley & Sons, Inc., New York, 464 pp.

Benjamin, J. R., and C. A. Cornell. 1970. Probability, Statistics, and Decision for Civil Engineers. McGraw-Hill, New York, 684 pp.

Blalock, H. M. Jr. 1972. Social Statistics. McGraw-Hill, New York, 583 pp.

Boneau, C. A. 1962. A Comparison of the Power of U and t Tests. Psychological Review, 69(3):246-256.

Box, G. E. P., W. G. Hunter, and J. S. Hunter. 1978. Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building. J. Wiley & Sons, Inc., New York, 653 pp.

Carlson, Robert E. 1977. A Trophic State Index for Lakes. Limnol. Oceanogr. 22(2):361-369.

Chapra, S. C. 1975. Comment on an Empirical Method of Estimating the Retention of Phosphorus in Lakes by W. B. Kirchner and P. J. Dillon. Water Resour. Res., 11(6):1033-1034.

Chapra, S. C., and K. H. Reckhow. 1979. Expressing the Phosphorus Loading Concept in Probabilistic Terms. J. Fish. Res. Board Can., 36(2):255-229.

Chatterjee, S. and B. Price. 1977. Regression Analysis by Example. J. Wiley & Sons, Inc., New York, 228 pp.

Cochran, W. G. 1963. Sampling Techniques. J. Wiley & Sons, Inc., New York.

Cornell, C. Allin. 1972. First-Order Analysis of Model and Parameter Uncertainty. In: Proceedings of the International Symposium on Uncertainties in Hydrologic and Water Resources Systems. University of Arizona, Tucson, Arizona. Vol. III, 1245-1274.

Dillon, P. J., and F. H. Rigler. 1975. A Test of a Simple Nutrient Budget Model Predicting the Phosphorus Concentration in Lake Water. J. Fish. Res. Board Can., 31(11):1711-1778.

Dobson, H. F. H., M. Gilbertson, and P. G. Sly. 1974. A Summary and Comparison of Nutrients and Related Water Quality in Lakes Erie, Ontario, Huron, and Superior. J. Fish. Res. Bd. Can. 31:731-738.

Duckstein, L., and I. Bogardi. 1978. Uncertainties in Lake Management. In: Proceedings of the International Symposium on Risk and Reliability in Water Resources. University of Waterloo, Waterloo, Ontario. pp. 638-661.

Edmondson, W. T. 1980. Secchi Disk and Chlorophyll. Limnol. Oceanogr. 25(2):378-379.

Freese, F. 1962. Elementary Forest Sampling. U.S. Dept. of Agriculture, Forest Service, Agriculture Handbook No. 232, 91 pp.

Hansen, M. H., W. N. Hurwitz, and W. G. Madow. 1953. Sample Survey Methods and Theory. J. Wiley & Sons, Inc., New York, 638 pp.

Hollander, M., and D. A. Wolfe. 1973. Nonparametric Statistical Methods. J. Wiley & Sons, Inc., New York, 503 pp.

Jessen, R. J. 1978. Statistical Survey Techniques. J. Wiley & Sons, Inc., New York, 520 pp.

Jones, J. R., and R. W. Bachmann. 1976. Prediction of Phosphorus and Chlorophyll Levels in Lakes. J. Water Pollut. Control Fed., 48(9):2176-2182.

Kleinbaum, D. G., and L. L. Kupper. 1978. Applied Regression Analysis and Other Multivariable Methods. Duxbury Press, North Scituate, Mass., 556 pp.

Larsen, D. P. 1980. personal communication.

Larsen, D. P., and H. T. Mercier. 1976. Phosphorus Retention Capacity of Lakes. J. Fish. Res. Board Can., 33(8):1742-1750.

Lee, David R. 1977. A Device for Measuring Seepage Flux in Lakes and Estuaries. Limnol. Oceanogr. 22(1):140-147.

Lorenzen, M. W. 1980. Use of Chlorophyll-Secchi Disc Relationships. Limnol. Oceanogr. 25(2):371-372.

Marsalek, J. 1975. Sampling Techniques in Urban Runoff Quality Studies. In: Water Quality Parameters, ASTM STP 573, American Society for Testing and Materials, pp. 526-542.

McGill, R., J. W. Tukey, and W. A. Larsen. 1978. Variations of Box Plots. Am. Stat. 32:12-16.

Megard, R. O., J. C. Settles, H. A. Boyer, and W. S. Combs, Jr. 1980. Light, Secchi Disks, and Trophic States. Limnol. Oceanogr. 25(2):373-377.

Montgomery, R. H., V. D. Lee, and K. H. Reckhow. 1980. A Comparison of Uncertainty Analysis Techniques: First Order Analysis vs. Monte Carlo Simulation. Paper presented at the International Association for Great Lakes Research Conference, Kingston, Ontario.

Mosteller, F., and R. E. K. Rourke. 1973. Sturdy Statistics: Nonparametric and Order Statistics. Addison-Wesley, Reading, Mass., 395 pp.

Mosteller, F., and J. W. Tukey. 1977. Data Analysis and Regression: A Second Course in Statistics. Addison-Wesley, Reading, Mass., 588 pp.

National Academy of Science and National Academy of Engineering. 1972. A Report of the Committee on Water Quality Criteria. Washington, D.C.

O'Hayre, A. P., and J. F. Dowd. 1978. Planning Methodology for Analysis and Management of Lake Eutrophication. Water Resources Bulletin 14(1):72-82.

Omernik, J. M. 1977. Nonpoint Source - Stream Nutrient Level Relationships: A Nationwide Study. U.S. Environmental Protection Agency, EPA-600/3-77-105. Corvallis, Oregon.

Popper, K. R. 1968. The Logic of Scientific Discovery. Harper Torchbooks, New York, 480 pp.

Porcella, D. B., S. A. Peterson, and D. P. Larsen. 1980. An Index to Evaluate Lake Restoration. Am. Soc. Civil Eng. Jour. (In Press).

Rast, W., and G. F. Lee. 1978. Summary Analysis of the North American (U.S. Portion) OECD Eutrophication Project: Nutrient Loading - Lake Response Relationships and Trophic State Indices. U.S. Environmental Protection Agency, EPA-600/3-79-008. Corvallis, Oregon.

Reckhow, K. H. 1977. Phosphorus Models for Lake Management. Ph.D. dissertation. Harvard Univ., Cambridge, Mass. 304 pp.

Reckhow, K. H. 1979a. Empirical Lake Models for Phosphorus: Development, Applications, Limitations, and Uncertainty, In: Perspectives on Lake Ecosystem Modeling. pp. 193-221. Edited by D. Scavia and A. Robertson. Ann Arbor Science Publishers, Ann Arbor, Mich.

Reckhow, K. H. 1979b. Quantitative Techniques for the Assessment of Lake Quality. U.S. Environmental Protection Agency, EPA-440/5-79-015. Washington, D.C. 146 pp.

Reckhow, K. H. 1979c. The Use of a Simple Model and Uncertainty Analysis in Lake Management. Water Resour. Bull, 15(3):601-611.

Reckhow, K. H. 1979d. Uncertainty Analysis Applied to Vollenweider's Phosphorus Loading Criterion. J. Water Pollut. Control Fed., 51(8):2123-2128.

Reckhow, K. H. 1979e. Sampling Designs for Lake Phosphorus Budgets. In: Proceedings of the Symposium on the Establishment of Water Quality Monitoring Programs, pp. 285-306. American Water Resources Association, Minneapolis, Minn.

Reckhow, K. H. 1980. Techniques for Exploring and Presenting Data Applied to Lake Phosphorus Concentration. Can. J. Fish. Aq. Sci. 37(2):290-294.

Reckhow, K. H., and Harbert Rice. 1975. 208 Modeling Approach. Report Prepared for the New Hampshire Lakes Region Planning Commission.

Reckhow, K. H., and S. C. Chapra. 1979. A Note on Error Analysis for a Phosphorus Retention Model. Water Resour. Res. 15(6):1643-1646.

Reckhow, K. H., and S. C. Chapra. 1980. Engineering Approaches for Lake Management: Data Analysis and Modeling. Ann Arbor Science, Ann Arbor, Mich. (In Press).

Reckhow, K. H., and J. T. Simpson. 1980. A Procedure Using Modeling and Error Analysis for the Prediction of Lake Phosphorus Concentration from Land Use Information. Can. J. Fish. Aq. Sci. 37(9):1439-1448.

Rechow, K. H., V. D. Lee, and S. C. Chapra. 1980a. An Examination of Lake Model Prediction Uncertainty Using First Order Analysis and Monte Carlo Simulation. Paper presented at the American Society of Limnology and Oceanography Annual Meeting. Los Angeles, Calif.

Reckhow, K. H., M. N. Beaulac, and J. T. Simpson. 1980b. Modeling Phosphorus Loading and Lake Response Under Uncertainty: A Manual and Compilation of Export Coefficients. U.S. Environmental Protection Agency, EPA-440/5-80-011. Washington, D.C. 214 pp.

Sakamoto, M. 1966. Primary Production by Phytoplankton Community in Some Japanese Lakes and Its Dependence on Lake Depth. Arch. Hydrobiol. 62:1-28.

Scavia, D. 1980. Uncertainty Analysis for a Lake Eutrophication Model. Ph.D. dissertation. University of Michigan, Ann Arbor, Mich.

Scavia, D., and A. Robertson, Eds. 1979. Perspectives on Lake Ecosystem Modeling. Ann Arbor Science Publishers, Inc., Ann Arbor, Mich.

Shannon, Earl E., and Patrick L. Brezonik. 1972. Eutrophication Analysis: A Multivariate Approach. Journ. San. Eng'g. Div. ASCE 98(1):37-57.

Simpson, J. T., and K. H. Reckhow. 1980. An Empirical Study of Factors Affecting Blue-Green versus Nonblue-Green Algal Dominance in Lakes. Office of Water Research and Technology, Dept. of Interior. Available from NTIS (PB 80169311).

Snedecor, G. W., and W. G. Cochran. 1967. Statistical Methods. Iowa State University Press, Ames, Iowa.

Snow, Phillip D., and Francis A. DiGiano. 1976. Mathematical Modeling of Phosphorus Exchange between Sediments and Overlying Water in Shallow Eutrophic Lakes. Report No. Env. E. 54-76-3. Dept. of Civil Engineering, University of Massachusetts. Amherst, Mass.

Thomann, R. V., D. M. DiToro, R. P. Winfield, and D. J. O'Connor. 1975. Mathematical Modeling of Phytoplankton in Lake Ontario, Part 1. Model Development and Verification. U.S. Environmental Protection Agency, EPA-660/3-75-005. Corvallis, Oregon.

Tukey, J. W. 1977. Exploratory Data Analysis. Addison-Wesley, Reading, Mass. 688 pp.

U.S. Environmental Protection Agency. 1974. The Relationships of Phosphorus and Nitrogen to the Trophic State of Northeast and North-Central Lakes and Reservoirs. National Eutrophication Survey Working Paper No. 23, USEPA, Corvallis, Oregon.

Uttormark, P. D., J. D. Chapin, and K. M. Green. 1974. Estimating Nutrient Loading of Lakes from Nonpoint Sources. U.S. Environmental Protection Agency, EPA-660/13-74-020. Corvallis, Oregon.

Vollenweider, R. A. 1968. The Scientific Basis of Lake and Stream Eutrophication, with Particular Reference to Phosphorus and Nitrogen as Eutrophication Factors, OECD (Organ. Econ. Coop. Dev.) Paris Tech. Rep. DAS/DSI/68.

Vollenweider, R. A. 1975. Input-Output Models With Special Reference to the Phosphorus Loading Concept in Limnology. Schweiz. Z. Hydrol., 37:53-84.

Vollenweider, R. A. 1976. Advances in Defining Critical Loading Levels for Phosphorus in Lake Eutrophication. Mem. Ist. Ital. Idrobiol., 33:53-83.

Walker, W. W., Jr. 1977. Some Analytical Methods Applied to Lake Water Quality Problems. Ph.D. dissertation. Harvard University, Cambridge, Mass. 528 pp.

Walker, W. W., Jr., 1979. Use of Hypolimnetic Oxygen Depletion Rate as a Trophic State Index for Lakes. Water Resour. Res. 15(6):1463-1470.

Williams, B. 1978. A Sampler on Sampling. J. Wiley & Sons, Inc., New York, 254 pp.

Wonnacott, T. H., and R. J. Wonnacott. 1972. Introductory Statistics. J. Wiley & Sons, Inc., New York, 510 pp.

☆ U.S. GOVERNMENT PRINTING OFFICE 1981 -757-064/0318