CHAPTER 4

Establishing an Appropriate Database

A. Introduction

The development of nationwide regional nutrient criteria requires the availability of an extensive amount of data from across the country for evaluation. Data may come from existing sources or can be collected from new sampling programs. Nutrient-related data for lakes and reservoirs, collected by various agencies for many different purposes, exist in various databases and have the potential to provide the basis for the development of nutrient criteria on a regional level. This chapter presents an overview of existing nutrient criteria databases and presents a general discussion on the evaluation of such data in terms of their use in the nutrient criteria development process. The chapter also provides a description of the process undertaken by EPA to use existing data from STORET and perhaps other existing data sets (e.g., U.S. Geological Survey [USGS] NAWQA) to generate preliminary nutrient criteria on an ecoregional level. In addition to discussing the use of existing data, the chapter discusses new data collection, including sampling design and the types of monitoring to be considered as part of data collection activities. The chapter ends with a general discussion of data management issues that are integral in the overall discussion of data storage and accessibility.

B. Evaluating Existing Data

In many States, historical data on lakes are extensive and may be sufficient to identify lake classes and reference lakes and to establish interim criteria. With societal interest in cultural eutrophication during the 1970's—programs such as EPA's Clean Lakes Program (CLP) and citizen's organizations such as LakeWatch— lakes have been extensively monitored and studied. Many historical data are available and may suffice for developing lake nutrient criteria. Although existing data may be sufficient, program managers should recognize that gathering, reducing, analyzing, and interpreting existing data sets can be costly and time consuming.

Development of nutrient criteria using the reference site approach requires a database that can be used to characterize (1) reference lakes and (2) the biological response of lakes to nutrient enrichment. At a minimum, observations should include total nitrogen (TN) (or Kjeldahl nitrogen plus NO_2-NO_3-N), total phosphorus (TP) (or total dissolved phosphorus plus particulate phosphorus), chlorophyll *a*, and Secchi depth. For each lake in the database, there should be information or inference on the status of anthropogenic nutrient loading to the lake. At a minimum, this information would be informed best professional judgment on whether anthropogenic nutrient loading was negligible or substantial for a given lake. The judgment could be based on personal observation, discharge information, land use information, or historical information. Such information may not reside together with the water quality observation and may need to be found and obtained separately.

As mentioned above, the water quality data should be sufficient to characterize a lake. All lakes in the database should have been sampled the same way and should be characterized the same. Ideally, lakes will have been sampled once during an index period that characterizes nutrient or trophic state (e.g., during spring overturn) and with sampling methods that are assumed to characterize the lake (e.g., pumped or composite sample of the entire water column). Some existing data sets may permit estimation of annual or growing season average nutrient concentrations.

Appropriate data analysis will be determined by the data set. The basic procedure is to consider each lake an independent sample unit and to estimate an annual characteristic value (annual average, median, minimum, maximum) of each water quality observation for each lake. These annual characteristic values are the information used to develop nutrient criteria and biological response of lakes to enrichment. This procedure assumes that lakes are independent and that annual averages among years are independent. The investigators and the Regional Technical Assistance Groups should decide whether the independence assumptions are reasonable and whether any modifications should be made.

1. Potential Data Sources

Databases could include water quality monitoring data from water quality agencies (often stored in STORET), national surveys such as the EPA Eastern Lakes Survey (Linthurst et al., 1986) or the Environmental Monitoring and Assessment Program (EMAP) (Paulsen et al., 1991), limnological studies, and volunteer monitoring information.

■ **STORET**

STORET is EPA's national data warehouse for water quality data. All State and Federal water quality monitoring agencies are required to submit their data to STORET. STORET is huge, covering all 50 States since the 1970's; it includes lakes, streams, rivers, and estuaries and physicochemical and biological data. STORET has no quality control for accepting or rejecting data, but it does require extensive metadata (data descriptors) that show how data were collected and what analytical methods were used. All sampling sites are referenced by latitude and longitude, by the EPA Reach File 3 (RF3), and by USGS hydrologic unit codes. Extracting data from STORET requires familiarity with the system as well as selection criteria (date ranges, location ranges, specified measured variables, collecting agency with known quality control) to keep from being overwhelmed with irrelevant data.

■ **National Eutrophication Survey (NES)**

EPA conducted NES in the early 1970's. Several hundred lakes were sampled and nutrient budgets were estimated. The lakes selected for the survey received discharges from municipal sewage treatment or the States requested they be included. NES contains a broad but incomplete sample of lakes, and therefore, the data from NES are not sufficient in themselves for developing reference conditions to support regional nutrient criteria. The NES data may be used for determining biological responses to enrichment and for developing site-specific criteria.

■ **National Surface Water Survey (NSWS)**

EPA conducted NSWS in the mid 1980's under the National Acid Precipitation Assessment Program. Lakes were surveyed only in those regions where they were initially thought to be at risk to acid precipitation: New England, the Adirondacks, the mid-Atlantic highlands, the mid-Atlantic coastal plain, the southeastern highlands (southern Appalachians and Ozark-Ouachitas), Florida, the upper Midwest, and the montane West. NSWS sampled 2,300 lakes ranging in size from 4 to 2,000 hectares; thus, the smallest and largest lakes were not represented. NSWS was a stratified random sample of lakes in the selected region; therefore, inferences can be made to the populations defined by the regions. Sampling took place in the fall. Nutrients were measured, so NSWS data could be used to help develop criteria.

■ Environmental Monitoring and Assessment Program

EPA's EMAP sampled lakes in New England and the Adirondacks in 1991, with a probability-based site selection procedure to obtain an unbiased sample (Paulsen et al., 1991). Further similar work is planned in the western States. The EMAP lake sample should include potential reference lakes as well as stressed lakes and could be used to identify reference lakes and characterize reference conditions.

■ Clean Lakes Program

The EPA CLP for restoring public lakes included a monitoring and assessment component. Lakes in this program were selected because they were perceived to have water quality impairment. Like data in the NES database, these data must be carefully scrutinized before being included in a database setting regional reference conditions to support nutrient criteria.

■ Volunteer Monitoring Programs

- Individual State Lake Association Programs that may contain considerable information, may be contacted through the National Directory of Environmental Monitoring Programs (contact: Alice Mayo, U.S. EPA headquarters, Washington, DC).

- National data on Secchi disc transparency have been collected since 1994 by the "Secchi Dip-In" (Carlson et al., 1997).

Elements of these databases could contribute to criteria development but, like the EMAP data, would need to be screened to identify reference and nonreference lakes.

■ State Monitoring Programs

Most States monitor some subset of lakes and impoundments within their borders for eutrophication and nutrient variables. Several of the more extensive lake monitoring programs (e.g., Minnesota, Wisconsin, Maine, Florida) are profiled in this document as examples of using monitoring data to help develop nutrient criteria. The purpose of the survey should be assessed before using the data. See Representativeness, in Section 2 below.

■ U.S. Army Corps of Engineers

The U.S. Army Corps of Engineers is responsible for more than 750 reservoirs. Extensive monitoring datahave been collected for many of these reservoirs that could contribute to the development of nutrient criteria for reservoirs.

■ U.S. Department of the Interior, Bureau of Reclamation (BuRec)

The Bureau of Reclamation manages many irrigation and water supply reservoirs in the West. Data from their operations may be available for some of these.

■ **Electric Utilities**

Many electric utilities own reservoirs for hydroelectric power generation, and the utilities are required to monitor the reservoirs' water quality. The largest of these, the Tennessee Valley Authority, has extensive chemical and biological monitoring data from most of its reservoirs from the early 1980's to the present.

2. Quality of Historical Data

The quality of older historical data sets is a recurrent problem because the data quality is often unknown. This is especially true of long-term repositories of data such as STORET and long-term State, academic, commission, or municipal databases, where objectives, methods, and investigators may have changed many times over the years. The most reliable data tend to be those collected by a single agency using the same protocol for a limited number of years. Supporting documentation should be examined to determine the consistency of sampling and analysis protocols.

When "mining" from large heterogeneous data repositories such as STORET, investigators must screen data for acceptance considering a number of variables, as discussed below.

■ **Location**

STORET data are georeferenced with latitude, longitude, and RF3 codes. These can be used to select specific locations or specific USGS hydrologic units. In addition, STORET often contains a site description. If selecting, for example, all lake sites within a geographic region, it is also important to know the rationale and methods of site selection by the original investigators. Such information may be included in STORET metadata, if known.

■ **Variables and Analytical Methods**

Thousands of variables are recorded in STORET records. Each separate analytical method yields a unique variable (called parameter in STORET); thus, five ways of measuring TP results in five unique variables. Because methods differ in accuracy, precision, and detection limits, it is generally unwise to mix methods in the same analysis. If there is one method that the investigator judges to be best, then only observations using that particular method can be selected. Selection of a particular "best" method may result in too few observations, in which case it may be more fruitful to select the most frequent method in the database. Some data may be missed because some methods may be synonymous, and there is an unknown component of error from incorrect data entry.

■ **Laboratory Quality Control**

Laboratory quality control data (blanks, spikes, replicates, known standards, etc.) are generally not reported in the larger data repositories. It is more cost-effective to accept or reject all data of the collecting agency or laboratory based on overall confidence of their quality control. Overzealousness in eliminating lower quality data can be counterproductive, because the increase in variance caused by analytical laboratory error may be negligible compared with natural variability or sampling error, especially for nutrients and related indicators.

■ **Collecting Agencies**

STORET data are identified by the agency that collected the data. Selecting only data from particular agencies with known, consistent collection and analytical methods and known quality will reduce variability due to unknown quality problems.

■ **Time Period**

Long-term records are critically important for establishing trends. In characterizing reference conditions for nutrient criteria, it is also important to determine if trends exist in the reference site database. For example, since passage of the Clean Water Act and elimination of most discharges to lakes, many lakes have improved markedly. Other lakes, subject to increased nonpoint-source runoff, may have declined in overall quality.

■ **Index Period**

If nutrient and water quality variables were measured more than once a year, an index period for estimating average concentrations must be set. The index period may be the entire year, spring or fall mixing (in regions where lakes stratify), or the summer growing season. The best index period is determined by investigators considering the characteristics of lakes of the region, the quality and quantity of data available, and estimates of temporal variability (if available).

■ **Representativeness**

Data may have been collected for specific purposes, such as developing nutrient budgets for eutrophic lakes. Such data are unlikely to be representative of the region or lake type of interest. The investigator must ask whether the lakes in the database are representative of the population of lakes to be characterized. If not, can a subset of representative lakes be selected from the database? If a sufficient sample of representative lakes (i.e., one large enough to characterize reference conditions) cannot be found, a new survey will be necessary (see section C below).

3. Data Reduction

To facilitate data manipulation and calculations, it is highly recommended that historical and present-day data be transferred to a relational database. Relational databases are powerful tools for data manipulation and initial data reduction (calculation of seasonal means, etc.). They allow selection of data by specific multiple criteria (e.g., all observations June–September in reference lakes of low alkalinity), calculation of means and totals by criteria, and definition and redefinition of linkages among data components.

Data reduction requires a clear idea of the analysis that will be attempted and a clear definition of the sample unit for the analysis. For example, a sample unit might be defined as “a lake basin during June–September.” For each variable measured, a mean value would then be estimated for each lake basin in each June–September index period on record. Analyses are then done with the observations (estimated means, medians, or modes) for each sample unit, not with the raw data. Steps in reducing the data include:

- Selecting the time period for analysis
- Selecting equivalent depths of sampling
- Selecting an index period to characterize lakes
- Selecting relevant chemical measures:
  -Quality of methods
  -Combining data from different methods
- Estimating values for analysis (mean, median, mode, minimum, maximum) based on the reduction selected

C. New Data Collection

When present-day and historical data do not exist or are inadequate for meeting desired objectives, collecting new monitoring data may be the only remaining alternative. A well-designed monitoring program is essential for establishing nutrient criteria in lakes. Monitoring data are needed for a variety of purposes, from the initial classification of lakes to assessing the effectiveness of controls. This section presents a brief backgound on statistical issues to consider when designing a monitoring program and provides recommendations for three types of monitoring associated with nutrient criteria.

Several manuals and statistics books are available that provide information on sampling design. The following manuals deal specifically with the sampling of lakes and reservoirs:

- Carlson, R., and J. Simpson. 1996. *A coordinator's guide to volunteer lake monitoring methods*. North American Lake Management Society. February 1996.
- Gaugush, R.F. 1986. *Statistical methods for reservoir water quality investigations*. Instruction Report E-86-2, U.S. Army Engineer Waterways Experiment Station, Vicksburg, Mississippi.
- Gaugush, R.F. 1987. *Sampling design for reservoir water quality investigations*. Instruction Report E-87-1, U.S. Army Engineer Waterways Experiment Station, Vicksburg, Mississippi.
- Reckhow, K.H. 1979. *Quantitative techniques for the assessment of lake quality*. U.S. EPA Office of Water Planning and Standards. EPA-440/5-79-015.
- Reckhow, K.H., and S.C. Chapra. 1983. *Engineering approaches for lake management. Volume 1: Data Analysis and Empirical Modeling*. Ann Arbor: Butterworth Publishers (Ann Arbor Science).

1. Types of Monitoring Associated With Nutrient Criteria

Monitoring is a critical component of nutrient criteria development and implementation. Monitoring is necessary after criteria are developed to determine trends and efficacy of water quality management in lakes and to improve the criteria with better information. If historical or existing data are not suitable for developing nutrient criteria, then a field survey is necessary to acquire the relevant data. More intensive diagnostic surveys and monitoring would be necessary if a lake did not meet its nutrient criteria, followed by long-term monitoring to determine if the management actions succeeded. Nutrient criteria thus are supported by three types of monitoring surveys: classification survey monitoring, diagnostic monitoring, and evaluation monitoring.

■ Classification Surveys to Support Nutrient Criteria Development

The purpose of a classification survey is to gather data that can be used to classify lakes or reservoirs into groups (classes) or along gradients with unique expected trophic status and trophic responses to enrichment. Classification depends on acquiring a database of lakes covering a gradient from least altered (reference lakes) to most altered in the region. The goal of the classification survey process is to identify classes of lakes independent of cultural eutrophication, that is, classes that would be recognizable if human cultural eutrophication were absent or minimal. Because the objective is to develop nutrient criteria, each lake class identified should have a unique suite of trophic conditions, as well as a unique response to cultural eutrophication, that sets each class apart from the other classes. Classes need not be discrete categories of lakes; classification also may identify natural environmental gradients rather than categories.

Existing regional lake databases potentially can be sufficient for classification in many regions of the country, as has been done in Minnesota (Heiskary et al., 1989) and Florida (Griffiths et al., 1997). However, if adequate data do not exist, new data should be collected.

Parameters to Survey

As many as possible of the water quality response parameters discussed in Chapter 5 (e.g., TP, TN, chlorophyll, Secchi depth, and dissolved oxygen) should be sampled during classification monitoring. In addition, classification also requires collecting information that helps explain observed patterns, including region (e.g., ecoregion or lake region); watershed size; lake water characteristics such as pH, conductivity, alkalinity, total suspended solids, and color; watershed land use; and human population density.

Sampling Frequency

Because the objective of classification monitoring is to characterize a large population of lakes, a single determination per lake may be the most cost-effective method (e.g., Linthurst et al., 1986). Multiple sample times (e.g., monthly) will yield a more precise estimate of annual average concentrations, but these must be weighed against the cost of sampling. For a given monitoring budget, it may be possible to sample 12 lakes 1 time for the same cost as sampling 1 lake 12 times. North temperate lakes or reservoirs are most effectively sampled during spring turnover for mass balance modeling. However, in general, ecoregional trophic state criteria will be indexed by EPA to the time period of optimal vegetative growth. This approach also is recommended for State and tribal sampling.

Sampling Location

Several alternatives are possible for sampling sites within lakes: single site (often the deepest point of the lake), multiple sites, spatially composite samples from multiple locations in a lake, or spatially composite samples from multiple sections of large lakes. Sampling location and frequency will affect not only the cost of the program but also the inferences that can be made from the sample. In general, composite samples are more cost-effective (information gained per dollar spent) than single sites or multiple locations.

To facilitate comparison, EPA recommends using the same spatial site location and discrete or compositing methods for both the classification survey and routine maintenance monitoring. If increased power is required for certain kinds of maintenance monitoring (see below), then the frequency of sampling should be increased, but the methodology should not be changed. Diagnostic investigation has entirely different objectives, so consistency with the classification sampling and maintenance monitoring is not necessary.

A single site may be chosen as the midpoint of the central basin of the lake, and this site may be sufficient to classify lakes within a region. The location of this single site should remain constant from one year to the next so that comparisons can be made during routine monitoring.

Composite samples are taken from several sites in a lake or lake zone and combined into a single sample for laboratory analysis. For example, water samples may be taken from four sites in a lake and poured into a single clean bucket. The composite sample is subsampled for chlorophyll *a* and nutrients. However, Secchi depth, temperature, and dissolved oxygen are measured at each of the four sites. Care must be taken that the methods and volume sampled are the same at each site. Composite samples characterize the lake better than a single sample, and they save laboratory costs. The principal disadvantage of composite samples is that they do not allow estimation of spatial variability within a lake.

Because large riverine reservoirs have known gradients of nutrients and productivity from the river inflow to the dam (Kennedy and Walker, 1990), a single site will not be appropriate. Large reservoirs would require a minimum of three sites, corresponding to the riverine, transitional, and lacustrine zones. Alternatively, and depending on reservoir design and size, sampling may be from the discharge or composited from the deepest portion of the reservoir. (See also Chapter 7, Section C, regarding a data reduction approach to ecoregional criterion.)

■ Diagnostic Monitoring

Diagnostic monitoring provides more detailed information on a specific lake or reservoir and allows the manager to identify key problems and develop an appropriate management plan. Diagnostic monitoring is carried out before and during lake restoration efforts, such as took place in the Clean Lakes Program. This process is more fully described in Chapter 8.

■ Evaluation Monitoring

The purpose of evaluation monitoring is to provide continuing information on the condition of a lake or reservoir. It can be used to determine if a lake is continuing to meet its nutrient criteria or to assess the effectiveness of any management controls that have been implemented. The level of effort required for evaluation monitoring is considerably less than for diagnostic monitoring. Examples of evaluation monitoring include operational lake monitoring by many State water quality agencies and volunteer monitoring efforts. Evaluation monitoring is discussed in more detail in Chapter 8.

Each of these monitoring objectives requires a statistically appropriate design to meet its objectives. The three types of monitoring are not entirely distinct, because data gathered for one purpose can be used for other purposes. Distinguishing between the different types of monitoring simply points out the need to identify ahead of time the purpose of the monitoring to maximize available resources.

2. Sampling Design

■ Specifying the Population and Sample Unit

Sampling is statistically expressed as a sample from a population of objects. In some cases, the population is finite, countable, and easy to specify—for example, all lakes in state X, where each lake is a single member of the population. In other cases, the population is more difficult to specify and may be infinite—for example, lake waters of State X, where any location in any lake defines a potential member of the population (Thompson, 1992). Sampling units may be natural units (entire lakes, cobbles in a littoral zone), or they may be arbitrary (plot, quadrate, sampling gear area, or volume) (Pielou, 1977). Finite populations may be sampled with corresponding natural sample units, but often the sample unit (e.g., a lake) is too large to measure in its entirety, and it must be characterized with one or more second-stage samples of the sampling gear (bottles, benthic grabs, quadrates, etc.)

Each sample unit is assumed to be independent of other sample units. The objective of sampling is to best characterize individual sample units to estimate some attributes (e.g., nutrient concentrations, dissolved oxygen) and their statistical parameters (e.g., mean, median, variance, percentiles) of a population of sample units. The objective of the analysis is to be able to say something about (estimate) the population. It is critical to distinguish between making an inference about a population of many lakes ("Reservoirs in the Blue Ridge are deep and oligotrophic") versus an inference about a single lake ("Lake has fewer fish species than unimpaired reference lakes"). These two kinds of inferences require different sampling designs: The first requires independent observations of many lakes and does not require repeated observations within sample units (pseudoreplication) (Hurlbert, 1984); the second inference often requires repeated observations within a lake. Examples of sample units include:

- A point in a lake (may be characterized by single or multiple sample device deployments). The population then would be all points in the lake, an infinite population.
- A constant area (e.g., square meter, hectare). The population could be all square meters of lake surface area in a State or region.
- A lake or a definable subbasin of a lake as a single unit. Because lakes are most often discrete environments, this is likely to be the most common sample unit. The population would be all lakes in a State or region, a finite population.

■ Specifying the Reporting Unit

It is also necessary to specify the units for which results will be reported. Usually, this is the population (e.g., all lakes), but often subpopulations (e.g., lakes within a given nutrient ecoregion) and even individual locations (e.g., lakes of special interest) will be specified. To help develop the sampling plan, it is useful to create hypothetical statements of results in the way that they will be reported, for example:

- Status of a place: "Lake X is degraded."
- Status of a region: "20% of the lake area in State X has elevated trophic state, above reference expectations"; "20% of lakes in State X have elevated trophic state."
- Trends at a place: "Nutrient concentrations in lake X have decreased by 20% since 1980."
- Trends of a region: "Average lake trophic state in State X has increased by 20% since 1980."; "Average trophic state index values in 20% of lakes of State X have increased by 15% or more since 1980."

- Relationships among variables: "50% increase of phosphorus loading above natural background is associated with decline in taxa richness of benthic macroinvertebrates, below reference expectations"; "Lakes receiving runoff from large impervious parking lots have 50% greater probability of elevated trophic state above reference conditions than lakes not receiving such runoff."

■ Sources of Variability

Variability of measurements has many possible sources. The intent of many sampling designs is to minimize the variability caused by uncontrolled or random effects and, conversely, to be able to characterize the variability caused by experimental or class effects. For example, lakes may be stratified by soil phosphorus content of the surrounding watersheds (e.g., Kiilsgaard et al., 1993) so that lakes within a soil phosphorus class may be likely to have similar water column TP concentrations. The population of lakes is stratified so that observations (sample units) from the same stratum will be more similar to each other than to sample units in other strata.

Environmental measures vary across different scales of space and time, and sampling design must consider the scales of variation. In lakes, measurements of some variable such as TP or chlorophyll concentrations are taken at single points in space and time (center of the lake, 2 m depth, 10 a.m. on 2 July). If the same measurement is taken at a different place (littoral zone, 1 m), lake, or time (30 January), the measured value may be different. A third component of variability is the ability to accurately measure the quantity in which we are interested, which can be affected by sampling gear, instrumentation, errors in proper adherence to field and laboratory protocols, and choice of methods used in making determinations.

The basic rule of efficient sampling and measurement is to sample so as to (1) minimize measurement errors; (2) maximize the components of variability that have influence on the central questions and reporting units; and (3) control other sources of variability that are not of interest, that is, minimize their effects on the observations. In the example of chlorophyll concentrations, variability could be reduced by sampling each of several lakes in the deepest part, with a vertically integrated pump sample taken in early spring before stratification appears. Many lakes are sampled to examine and characterize the variability due to different lakes (the sampling unit). Each lake is sampled in the same way, in the same place, and in the same timeframe to minimize variability caused by location, depth, and season, which are not of interest in this particular study.

In the above example, chlorophyll concentrations vary with location within a lake and among lakes and with time of sampling (day, season, year). If the spatial and temporal components of variability within lakes are large (e.g., measurements of chlorophyll concentrations typically vary more between spring and fall samples within a lake than they do among lakes), then it may be best either to use an index period sample or to estimate a composite from several determinations. For this reason, lake chlorophyll concentrations often are estimated as a growing season average, estimated from several determinations (e.g., monthly) during the growing season.

In statistical terminology, there is a distinction between sampling error and measurement error that has little to do with actual errors in measurement. Sampling error is the error attributable to selecting a certain sample unit (e.g., a lake or a location within a lake) that may not be representative of the population of sample units. Statistical measurement error is the ability of the investigator to accurately characterize the sampling unit. Thus, measurement error includes components of natural spatial and temporal variability within the sample unit as well as actual errors of omission or commission by the investigator. Measurement error is minimized with methodological standardization: selection of cost-effective low variability sampling methods, proper training of personnel, and quality assurance procedures to minimize methodological errors. In analytical laboratory procedures, measurement error is estimated by replicate determinations on some subset of samples (but not necessarily all). Similarly, in field investigations, some subset of sample units should be measured more than once to estimate measurement error.

Analysis of variance can be used to estimate measurement error. All multiple observations of a variable are used (from all lakes with multiple observations), and lakes are the primary effect variable. The root means square error (RMSE) of the analysis of variance is the estimated variance of repeated observations within lakes. Note that a hypothesis test (F-test) is not of interest in this application, only the RMSE of the analysis.

Natural variability that is not of interest for the questions being asked, but that may affect the ability to address them, should be estimated with the RMSE method above. If the variance estimated from RMSE is unacceptably large (i.e., as large or larger than variance expected among sample units), then it is often necessary to alter the sampling protocol, usually by increasing sampling effort in some way, to further reduce the measurement error. Measurement error can be reduced by multiple observations at each sample unit, for example, multiple ponar casts at each sampling event, multiple observations in time during a growing season or index period, depth-integrated samples, or spatially integrated samples.

A less costly alternative to multiple measures in space is spatially composite determinations. In nutrient or chlorophyll determinations, a water-column-pumped sample, where the pump hose is lowered through the water column, is an example of a spatially composite determination. Spatially integrating an observation and compositing the material into a single sample is almost always more cost-effective than retaining separate multiple observations. This is especially true for relatively costly laboratory analyses such as organic contaminants and benthic macroinvertebrates.

Statistical power is the ability of a given hypothesis test to detect an effect that actually exists, and it must be considered when designing a sampling program (e.g., Peterman, 1990; Fairweather, 1991). The power of a test ($1-b$) is defined as the probability of correctly rejecting the null hypothesis (H0) when H0 is false (i.e., the probability of correctly finding a difference [impairment] when one exists). For a fixed confidence level (e.g., 90%), power can be increased by increasing the sample size or the number of replicates. To evaluate power and determine sampling effort, an ecologically meaningful amount of change in a variable must be set.

Optimizing sampling design requires consideration of tradeoffs among the measures used, the effect size that is considered meaningful, desired power, desired confidence, and resources available for the sampling program. Every study requires some level of repeated measurement of sampling units to estimate precision and measurement error. Repeated measurement at 10% or more of sites is common among many monitoring programs.

■ Alternative Sampling Designs

Sampling design is the selection of a part of a population to observe the attributes of interest. To estimate the values of those attributes for the whole population, classical sampling design makes assumptions about the variables of interest. In particular, it assumes that the values are fixed (but unknown) for each member of the population until that member is observed (Thompson, 1992). This assumption is perfectly reasonable for some variables —for example, length, weight, and sex of members of an animal population—but it seems less reasonable for more dynamic variables such as nutrient concentrations, loadings, or chlorophyll concentrations of lakes. Designs that assume that the observed variables are themselves random variables are model-based designs, in which previous knowledge or assumptions are used to select sample units.

Probability-Based Designs (random sampling)

The most basic probability-based design is simple random sampling, in which all possible sample units in the population have the same probability of being selected, that is, all possible combinations of n sample units have equal probability of selection from among the N units in the population. If the population N is finite and not excessively large, a list can be made of the N units, and a sample of n units is randomly selected from the list. This is termed list frame sampling. If the population is very large or infinite (such as locations in a lake), one can select a set of n random (x,y) coordinates for the sample.

All sample combinations are equally likely in simple random sampling; thus, there is no assurance that the sample actually selected will be representative of the population. Other unbiased sampling designs that attempt to acquire a more representative sample include stratified, systematic, and multistage designs. In stratified sampling, the population is subdivided or partitioned into strata, and each stratum is sampled separately. Partitioning is typically done so as to make each stratum more homogeneous than the overall population; for example, lakes could be stratified on ecoregion. Systematic sampling is the systematic selection of every kth unit of the population from one or more randomly selected starting units, and it ensures that samples are not clumped in one region of the sample space. Multistage sampling requires selection of a sample of primary units, such as fields or hydrologic units, and then selection of secondary sample units, such as plots or lakes within each primary unit in the first-stage sample.

Estimation of statistical parameters requires weighting of the data with inclusion probabilities (the probability that a given unit of the population will be in the sample) specified in the sampling design. In simple random sampling, inclusion probabilities are by definition equal, and no corrections are necessary. Stratified sampling requires weighting by the inclusion probabilities of each stratum. Unbiased estimators have been developed for specific sampling designs and can be found in sampling textbooks, such as Thompson (1992).

Model-Based Designs

Use of probability-based sampling designs may miss relationships among variables (models), especially if there is a regression-type relationship between an explanatory and a response variable. As an example, elucidation of lake response to phosphorus loading with the Vollenweider model (Vollenweider, 1968) required a range of trophic states from ultraoligotrophic to hypereutrophic. A simple random sample of lakes is not likely to capture the entire range (i.e., there would be a large cluster of mesotrophic lakes with few at high or low ends of the trophic scale), and the random sample may therefore be biased with respect to the model.

In model-based designs, sites are selected based on previous knowledge of auxiliary variables, such as estimated phosphorus loading, lake depth, and elevation. Often, these designs preclude an unbiased estimate of the population response variable (e.g., trophic state), unless the model can be demonstrated to be robust and predictive, in which case the population value is predicted from the model and from previous knowledge of the auxiliary (predictive) variables. Selection of unimpacted reference sites is an example of samples for a model (index development; response of index variables to measures of anthropogenic influence) that cannot later be used for unbiased estimation of the biological status of lakes. Ideally, it may be possible to specify a design that allows unbiased estimation of both population and model. Statisticians should be consulted in developing the sample design for a nutrient criteria and monitoring program.

D. Database Management

Critical to the success of any monitoring and criteria development program is comprehensive data management. Because agencies that engage in water quality management are required to be accountable, because monitoring and assessment tools continue to be developed, and because desktop computers are now capable of many of the same tasks as larger computers, data management must be addressed at the outset of a program or project. Storing data in filing cabinets or on spreadsheet files is no longer adequate.

The most powerful database architecture for storing large, complex data sets with multiple relationships among the data elements are relational databases. The hierarchical nature of a watershed and a survey and assessment program are reflected in a relational database: a watershed may contain many sampling sites; each site may be sampled multiple times during an investigation; and each sample may be tested for multiple constituents.

Nonrelational databases consist of one or more flat files. Examples include text files and spreadsheet files. Flat file databases have a number of disadvantages, including redundant data, maintenance difficulties, and slow access to data. A number of relational database management systems (i.e., software) exist today, including Oracle, Sybase, DB2, Informix, SQL server, and ACCESS.

In the simplest form, a relational database begins as a collection of tables. Each table is related to at least one other table through one or more key fields that act as links (i.e., relationships) between tables. Because tables are related, information can be retrieved from more than one table at a time. Tables generally contain data about a particular subject. For instance, data about a station that is sampled would be in one table, whereas data about a sampling event would be in a separate but related table. Good database design includes reducing or eliminating duplicate information in tables and being able to make changes to the database tables as data needs change. Data should be easy to manage, aggregate, retrieve, and analyze. EPA has sponsored the development of two relational databases that are available for data management for nutrient criteria, modernized STORET and Ecological Data Application System (EDAS).

1. Modernized STORET

EPA's Office of Water has redesigned and modernized STORET. The historical data contained in STORET, BIOS, and ODES are available in modernized STORET. This database was designed to meet emerging data and information needs associated with watershed level environmental protection. The features of the new system were carefully engineered to meet the information requirements of Federal, State, and local clients engaged in ambient water quality and biological monitoring activities of all kinds. Modernized STORET meets the following five requirements set by EPA:

- It must be easy to enter and retrieve data, and the system must be web enabled.
- The system must have menu access and browse capability.

- The system must support the storage of quality assurance and quality control information on a project and result basis.

- The system must be flexible and able to change with the changing needs of its users.

- The system must provide a wide range of standard output forms, including Geographic Information System environments.

STORET centrally stores all data submitted by all agencies, and it allows each State or agency to enter, store, and use its own data. It is a flexible and generalized data warehouse, both within-agency and nationwide, and as such does not have program-specific analysis capabilities built in.

2. EDAS

EDAS is EPA's biological metric calculation software. It is entirely compatible with STORET, and final data should reside in STORET. EDAS contains built-in data reduction and recalculation queries that are used in biological assessment. It is designed to enable the user to easily manage, aggregate, integrate, and analyze data to make informed decisions regarding the condition of a water resource. Biological assessment and monitoring programs require aggregation of raw biological data (lists and enumeration of taxa in a sample) into informative indicators. EDAS is designed to facilitate data analysis, particularly the calculation of biological metrics and indexes. Predesigned queries that calculate a wide selection of biological metrics are included with EDAS.