2. DEVELOPING A MONITORING PLAN

2.1 INTRODUCTION

Since the relationship between public health and water quality began to influence legislation in the early 1900s, water quality management and its related information needs have evolved considerably. Today, the Intergovernmental Task Force on Monitoring Water Quality (ITFM, 1995b) defines water quality monitoring as an integrated activity for evaluating the physical, chemical, and biological character of water in relation to human health, ecological conditions, and designated water uses. Water quality monitoring for nonpoint sources of pollution includes the important element of relating the physical, chemical, and biological characteristics of receiving waters to land use characteristics. Without current information, water quality and the effects of land-based activities on water quality cannot be assessed, effective management and remediation programs cannot be implemented, and program success cannot be evaluated.

The most fundamental step in the development of a monitoring plan is to define the goals and objectives, or purpose, of the monitoring program. In the past, numerous monitoring programs did not document this aspect of the design process and the resulting data collection efforts led to little useful information for decision making (GAO, 1986; MacDonald et al., 1991; National Research Council, 1986; Ward et al., 1990). As a result, the identification of monitoring goals is the first component of the design framework outlined by the ITFM (1995b). In general, monitoring goals are broad statements such as “to measure improvements in Elephant Butte Reservoir” or “to verify nutrient load reductions into the Chesapeake Bay.” Designing a monitoring plan also includes selecting sampling variables, a sampling strategy, station locations, data analysis techniques, the length of the monitoring program, and the overall level of effort to be invested. Figure 2-1 presents one approach for developing a monitoring plan.

Monitoring programs can be grouped according to the following general purposes or expectations (ITFM, 1995b; MacDonald et al., 1991):

- Describing status and trends
- Describing and ranking existing and emerging problems
- Designing management and regulatory programs
- Evaluating program effectiveness
- Responding to emergencies
- Describing the implementation of best management practices
- Validating a proposed water quality model
- Performing research

The remainder of the design framework outlined by the ITFM (1995b) includes coordination and collaboration, design, implementation, interpretation, evaluation of the monitoring program, and communication. Numerous guidance documents have been developed, or are in development, to assist resource managers in developing and implementing monitoring programs that address all aspects of the ITFM's design framework. Appendix A presents a review of more than 40 monitoring guidances for both point and nonpoint source pollution. These guidances discuss virtually every aspect of nonpoint source pollution monitoring, including monitoring program design and objectives, sample types and sampling methods, chemical and physical water quality variables, biological monitoring, data analysis and management, and quality assurance and quality control.

Once the monitoring goals have been established, existing data and constraints should be considered. A thorough review of literature pertaining to water quality studies previously conducted in the geographic region of interest should be completed before starting a new study. The review should help determine whether existing data provide sufficient information to address the monitoring goals and what data gaps exist.

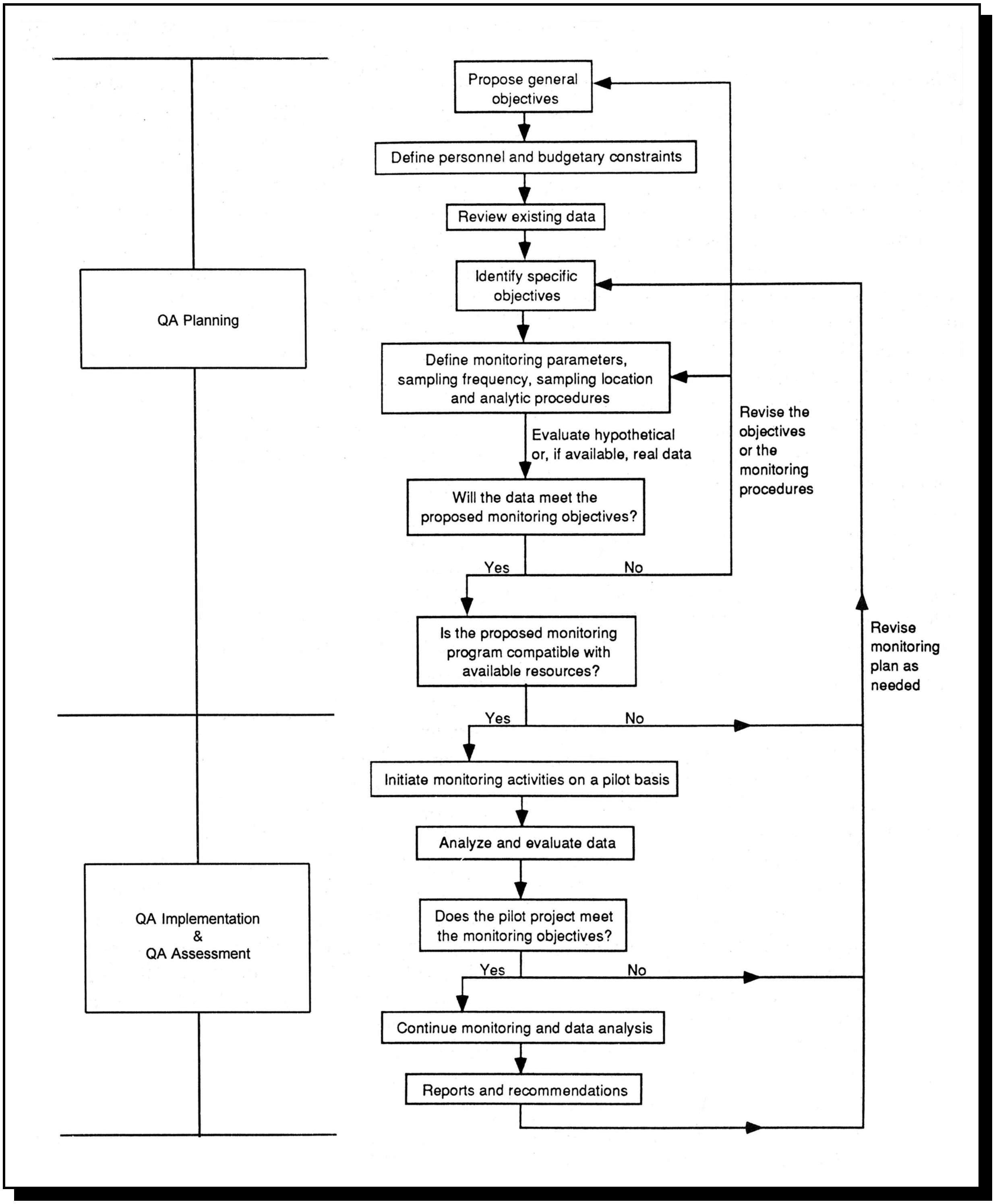


Figure 2-1. Development of a monitoring project (after MacDonald et al., 1991).

Identification of project constraints should address financial, staffing, and temporal elements. Clear and detailed information should be obtained on the time frame within which management decisions need to be made, the amounts and types of data that must be collected, the level of effort required to collect the necessary data, and the equipment and personnel needed to conduct the monitoring. From this information it can be determined whether available personnel and budget are sufficient to implement or expand the monitoring program.

As with monitoring program design, the level of monitoring that will be conducted is largely determined when goals and objectives are set for a monitoring program, although there is some flexibility for achieving most monitoring objectives. Table 2-1 provides a summary of general characteristics of various types of monitoring.

The overall scale of a monitoring program has two components—a temporal scale and a geographic scale. The temporal scale is the amount of time required to accomplish the program objectives. It can vary from an afternoon to many years. The geographic scale can also vary from quite small, such as plots along a single stream reach, to very large, such as an entire river basin. The temporal and geographic scales, like a program's design and monitoring level, are primarily determined by the program's objectives. Hence, unspecific or unclear monitoring objectives present a barrier to selecting the appropriate temporal and geographic scales.

If the main objective is to determine the current biological condition of a stream, sampling at a few stations in a stream reach over 1 or 2 days might suffice. Similarly, if the monitoring objective is to determine the presence or absence of a nonpoint source impact, a synoptic survey might be conducted in a few select locations. If the objective is to determine the effectiveness of a nutrient management program for reducing nutrient inputs to a downstream lake, however, monitoring a subwatershed for 5 years or longer might be necessary. If the objective is to calibrate or verify a model, more intensive sampling might be necessary.

Depending on the objectives of the monitoring program, it might be necessary to monitor only the waterbody with the water quality problem or it might be necessary to include areas that have contributed to the problem in the past, areas containing suspected sources of the problem, or a combination of these areas. A monitoring program conducted on a watershed scale must include a decision about a watershed's size. The effective size of a watershed is influenced by drainage patterns, stream order, stream permanence, climate, number of landowners in the area, homogeneity of land uses, watershed geology, and geomorphology. Each factor is important because each has an influence on stream characteristics, although no direct relationship exists.

There is no formula for determining appropriate geographic and temporal scales for any particular monitoring program. Rather, once the objectives of the monitoring program have been determined, a combined analysis of them and any background information on the water quality problem being addressed should make it clear what overall monitoring scale is necessary to reach the objectives.

Other factors that should be considered to determine appropriate temporal and geographic scales include the type of water resource being monitored and the complexity of the nonpoint source problem. Some of the constraints mentioned earlier, such as the availability of resources (staff and money) and the time frame within which managers require monitoring information, will also contribute to determination of the scales of the monitoring program.

Table 2-1. General characteristics of monitoring types .

Type of Monitoring	Number and Type of Water Quality Parameters	Frequency of Measurements	Duration of Monitoring	Intensity of Data Analysis
Trend	Usually water column	Low	Long	Low to moderate
Baseline	Variable	Low	Short to medium	Low to moderate
Implementation	None	Variable	Duration of project	Low
Effectiveness	Near activity	Medium to high	Usually short to medium	Medium
Project	Variable	Medium to high	Greater than project duration	Medium
Validation	Few	High	Usually medium to long	High
Compliance	Few	Variable	Dependent on project	Moderate to high

Source: MacDonald et al., 1991.

2.2 MONITORING OBJECTIVES

Identifying and concisely stating the monitoring objectives are critical steps in the development of a monitoring program. Unlike monitoring goals, monitoring objectives are more specific statements that can be used to complete the monitoring design process including scale, variable selection, methods, and sample size (Plafkin et al., 1989; USDA-NRCS, 1996). Monitoring program objectives must be detailed enough to allow the designer to define precisely what data will be gathered and how the resulting information will be used. Vague or inaccurate statements of objectives lead to program designs that provide too little or too much data, thereby failing to meet management needs or costing too much.

Monitoring programs can be implemented for one or many reasons. The more common types of monitoring program objectives are summarized below. The emphasis of this guidance is on evaluation monitoring, but information contained herein might also be used to address other types of monitoring. The reader is cautioned that even though two different monitoring programs might share some objective listed below, their designs can be radically different.

2.2.1 Monitoring Objective Category: Problem Definition

(1) Determine whether an impairment exists

Meeting this objective involves an investigation of key parameters to determine the general condition of a habitat or water quality. Measurements of individual pollutants in waterbodies are often taken to determine whether violations of water quality standards are occurring. Biological monitoring is also useful when evaluating whether designated uses are supported. Monitoring associated with this type of objective might reveal that a suspected problem is more complicated or serious than originally thought and that more intensive monitoring studies will be necessary.

(2) Determine the extent of the impairment

Even if a problem is known to exist, the geographic and temporal extent of the problem might not be known. Does the problem affect a stream reach, or does the problem extend to the downstream lake? Some pollution sources are emitted only during certain parts of the year or in association with certain events, such as storms, or might be a problem only during a particular time of the year, such as fish spawning season. Determining the geographic and temporal aspects of a pollution problem will help focus management on BMP systems that will have the most benefit.

(3) Determine the causes and sources of impairment

Monitoring might be required to determine the cause of an environmental problem, such as degraded fish habitat or an algal bloom. Determining the pollution's source is often more difficult than determining its presence because there are often many potential sources whose influences overlap. When conducting monitoring for this purpose, it is important to monitor the appropriate water quality characteristics and account for climatic factors to establish a cause-and-effect relationship, even though it might be difficult to prove.

Point and nonpoint sources often affect the same waterbody, and monitoring might also be required to determine the contribution and relative importance of each to water quality impairment. It might also be necessary to determine which areas are the most critical in causing waterbody impairment. For instance, a high erosion rate on land far from a receiving waterbody might have a lower pollution-causing potential than an area with a lower erosion rate near to a receiving waterbody. Factors such as the timing of pollutant contributions relative to the hydrologic cycle of the waterbody and the ecology of the biological communities must be factored into the analysis. In addition, the distance of pollutant sources from receiving waters, the fate and transport of pollutants from different sources, the magnitude of pollutant contributions from each source, and the distance to the impaired resource of concern (as distinguished from distance to a point of entry into a receiving waterbody, which might be some distance from the actual impairment) should be considered. This type of information can often be used in developing load allocations for nonpoint pollution sources and wasteload allocations for point sources, although extensive monitoring might be required.

2.2.2 Monitoring Objective Category: Model Development

(1) Calibrate models

Model calibration is the first stage of testing a model and tuning it to a set of field data. Field data are necessary to guide the modeler in choosing the empirical coefficients in a model before the model can be used to predict the effect of management techniques or activities.

(2) Validate models

Model validation involves the testing of a model using a second set of field data. In most cases, the second set of field data should represent an independent data set that extends the range of conditions for which the model is valid. If an independent data set is not available, a set of randomly selected data should be used for validation. Once a model has been validated, it can be used to assist managers with management decisions within the range of the calibration and validation data sets.

2.2.3 Monitoring Objective Category: Evaluation (emphasis of this guidance)

(1) Measure the effectiveness of best management practice (BMP) systems

Individual BMPs or groups of BMPs are monitored to determine the extent of pollution control. Monitoring for individual BMPs can typically be conducted at a plot or field scale, whereas monitoring for BMP systems is usually conducted on a watershed scale because the combined effect of a few or several BMPs is being investigated. Studies of some individual practices can be conducted in a relatively short time (less than 5 years), while others might take longer. Evaluation of BMP systems is typically conducted over a long term (more than 5 years) because BMP implementation can take years to affect water quality. This type of monitoring is difficult due to the presence of pollutant reserves in soil and sediments, the effect of many land uses within a study area, the variety of approaches that landowners use to implement similar systems of BMPs, and the need to track land management as well as water quality and climatic variables.

(2) Analyze trends

The objective here is to answer the question, “Is water quality changing over time?” Baseline monitoring is part of trend analysis because establishing a baseline is essential to analyzing trends. However, baseline monitoring is generally thought of as determining a condition prior to pollutant entry or prior to a change in waterbody condition, whether beneficial or detrimental. Controlling for influencing factors such as climate is necessary if baseline monitoring is to be used as a reference point for trend analysis and management decisions. The ability to relate water quality changes to changes in land management depends on the quality and quantity of data collected on land management practices.

2.2.4 Monitoring Objective Category: Conduct Research

Research monitoring is done to address specific research questions. Research monitoring is usually conducted on a plot scale, is well controlled, and is limited to a very specific question. Monitoring and data analysis techniques for research and for other types of monitoring are often very similar, and the difference between them is often one of objective rather than approach. A critical examination of articles about relevant and well-conducted research projects in which monitoring is a key element can provide excellent guidance for the design of a monitoring program.

2.3 DATA ANALYSIS AND PRESENTATION PLANS

Ward et al. (1990) point out that one of the most important and difficult tasks is to identify what information is to be produced by the monitoring effort. It is particularly critical to ensure that policy makers and other stakeholders know the type of information that a monitoring program can produce and that realistic monitoring program expectations are developed. Ward et al. (1990) identify key steps to ensure that realistic expectations are placed on the monitoring program and the associated data analysis:

- Perform a thorough review of the legal basis for the management effort and define the resulting “implications” for monitoring.

- Review the administrative structure and procedures developed from the law in order to define the information expectations of the management staff.

- Review the ability of the monitoring program to supply information.

- Formulate an information expectations report for the monitoring system.

- Present the information expectations report to all users of the information.

- Develop consensus as to an agreeable formulation of information expectations and related monitoring system design criteria.

This process is typically performed as an iterative process that involves the technical staff and the decision makers who developed the monitoring objectives. To develop an information expectations report, the data analyst might need to have formal meetings, develop questionnaires, and conduct interviews to learn what the managers need. In some cases this iterative process might require modifying or redesigning the monitoring program. The data analyst should remember that complete consensus might not be possible.

When developing an information expectations report, the presentation of results should be selected depending on the audience reviewing the information and the objectives of the monitoring program. How quickly must information be presented to information users? To what kind of information and how much information do the decision makers respond favorably? At a minimum, the data analyst should prepare example report formats to be approved by the decision makers, keeping in mind that “a picture is worth a thousand words.” In all cases, the goal should be to present clear and accurate information that is not subject to misinterpre-tation. Ward et al. (1990) present an example outline (Figure 2-2) of what might be considered in an expectations report. (The data analyst should modify this outline to suit individual needs.)

2.4 VARIABLE SELECTION

In these days of increasing monitoring and evaluation needs and relatively small monitoring and evaluation budgets, it is extremely important for program managers to design efficient monitoring and evaluation programs. The variables selected for a monitoring program should be tied directly to the monitoring objectives. It is often the case that some variables in addition to those of prime interest are monitored because they are relatively cheap to monitor and might provide some useful information for purposes not yet outlined. This is generally reasonable, but the technical staff should (1) anticipate these undefined purposes so that the extra variables are monitored in a manner that yields useful information (e.g., support statistical analyses) and (2) make sure the extra cost associated with monitoring additional variables does not preclude necessary expansions or extensions of the monitoring and evaluation program for the variables of prime interest.

In many instances the water quality problem will directly indicate what variables should be monitored. For example, a dissolved oxygen problem would strongly suggest monitoring of dissolved oxygen. (Typically, biochemical oxygen demand, sediment oxygen demand, temperature, and nutrients would be monitored as well.) Or, if the goal is to assess the impact of nonpoint source controls in terms of standards violations, then the variables selected should be those required for the analysis of standards violations.

Expectations Report Outline

- Evolution of Water Quality Management Program
    - Geographical/Hydrological Setting
    - Water Quality Problems
    - Water Quality Laws
    - Management Program Structure
    - Management Procedures

- Information “Expected” by Management Program
    - Implications of the Law Establishing the Program
    - Legal Goals
    - Management Powers and Functions
    - Monitoring Requirements Directly Stated
    - Information Needs of Management Operations
    - Water Quality Criteria
    - Water Quality Standards
    - Permits
    - Compliance
    - Enforcement
    - Construction Loans
    - Planning
    - Water Quality Assessment

- Ability of Monitoring Systems to Produce Water Quality Information
    - Narrative Information
    - Numerical Information-Data
    - Graphical Information
    - Statistical Information
    - Average Conditions
    - Changing Conditions
    - Extreme Conditions
    - Water Quality Indices

- Suggested Information Expectations for Monitoring System
    - Management Information Goal(s)
    - Definition of Water Quality
    - Monitoring System Goal(s)
    - Information Product of Monitoring System
    - Narrative
    - Graphical
    - Statistical

- Resulting Monitoring Network Design Criteria
    - Variable Selection
    - Site Selection
    - Sampling Frequency Determination

Figure 2-2. Expectations report outline (Ward et al., 1990).

In some cases, it might be more beneficial to use surrogate measures instead of the variables mentioned in the monitoring goals and objectives. In these cases, objectives for the surrogates that are consistent with the overall monitoring and evaluation goals should be established. The key to using surrogate measures is to be certain that a reliable relationship exists between the true measure and the surrogate measure. For example, if the objective is to monitor the condition of salmon spawning areas, surrogate measures are necessary because the condition of salmon spawning areas is a composite of many factors. Good surrogate variables would be stream bank undercut, embeddedness, and vegetative overhang (Platts et al., 1983). The corresponding surrogate goals could be to reduce cobble embeddedness and to increase vegetative overhang to appropriate levels for salmon spawning. The monitoring goals would then be to document changes in cobble embeddedness and vegetative overhang.

Poor surrogate selection results when a known relationship between the monitoring goals and objectives and the chosen surrogate measures does not exist. For example, a poor surrogate for estimators of sediment delivery to water resources is the unqualified use of erosion rates. Without the existence of a known relationship between these two measures (i.e., sediment delivery ratio), the surrogate will produce misleading results.

Variable selection should also reflect the nonpoint source data analysis and presentation plan. For example, if the plan involves data normalization or grouping prior to data analysis, the variable list should include those variables used to normalize and/or group the data. Some analyses might require discrete observations, whereas others might use continuous data. All monitoring sites should be characterized sufficiently for meaningful data interpretation, including georeferencing. For surface water sites the relevant information may include waterbody name, river reach number and milepoint, location, prevailing winds, shading, bottom sediment, elevation, slope, stream width and depth, drainage area, upstream land use, lake depth, and more. In the case of ground water monitoring, this information includes the aquifer tapped by a well, the depth of the well, the type of well construction, and the well elevation (USGS, 1977). Water level measurements should be included in all ground water studies.

Since there are numerous variables to choose from but monitoring budgets are limited, some method to prioritize variable selection is often necessary. When available, existing data should be used to guide variable selection. Further discussion on variable selection, prioritization, and optimization are provided by USDA-NRCS (1996), MacDonald et al. (1991), and Sherwani and Moreau (1975). In some cases, optimal variable selection is not possible, perhaps due to lack of local data. In such cases, the researcher might need to rely on professional judgment and the review of monitoring programs of similar nature and scope.

Some data requirements for nonpoint source monitoring and evaluation efforts can be met using nationally available data sources. Appendix B describes some of these data sources and includes information for those interested in accessing the data. Other data sources are available to nonpoint source professionals as well, and state, regional, or local sources of data in particular should be investigated. State agriculture, forestry, and other environmental agencies; counties; municipalities; and state and local health departments are likely sources of water quality, health-related, and land use data and information. Regional planning commissions, local universities, and environmental consultants might also be able to provide data. The sources summarized in Appendix B focus on the major data sources made available to EPA or known to reviewers of this document. The remainder of this section summarizes key data that would normally be considered in a nonpoint source monitoring program.

2.4.1 Physical and Chemical Water Quality Data

Physical and chemical water quality data are essential to almost all nonpoint source monitoring and evaluation efforts, due to the relationships between flow and pollutant characteristics. For example, it might be necessary to establish watershed water budgets so that the location and magnitude of nonpoint sources or background sources can be determined. In other cases, the extent of the floodplain might prove critical to assessments of BMP control needs. Important physical and chemical water quality variables to monitor include flow (streams), temperature, transparency, suspended sediment, sedimentation rate, dissolved oxygen, pH, conductivity, alkalinity/acid neutralizing capacity (lakes), and nutrients. Other factors, such as cobble embeddedness, woody debris, and salinity, might be important depending on type of water body and monitoring goals.

2.4.2 Biological Data

Biological data can be very useful for evaluating water resource impairment due to nonpoint source impacts because aquatic organisms integrate the exposure to various nonpoint sources over time. Measures of biological communities integrate the effects of different pollutant stressors—excess nutrients, toxic chemicals, increased temperature, excessive sediment loading, and others—and thus provide an overall measure of the aggregate impact of the stressors. Monitoring changes in aquatic communities over time can serve as a measure of improvement due to BMPs. The biological survey approach used depends on waterbody type, i.e., stream, river, lake, wetland, or estuary. Important biological parameters to monitor include bacteria, algal biomass, macrophyte biomass and location, macroinvertebrates, and fish populations.

2.4.3 Precipitation Data

Precipitation data, including total rainfall, rainfall intensity, storm interval, and storm duration, have proven to be key to successful interpretation of nonpoint source data in the Nationwide Urban Runoff Program (NURP), Model Implementation Program (MIP), and Rural Clean Water Program (RCWP) studies. By combining precipitation data with pollutant loading evaluations, it has been found that a few storms can account for a large proportion of the total annual pollutant load. Johengen and Beeton (1992) found that, in the Saline Valley RCWP, a few storms accounted for more than 50 percent of the annual loading. Interestingly, they found that initial estimates of suspended solids and phosphorus loadings were only 20 and 50 percent of loadings estimated by adjusting for daily precipitation. The project-mandated weekly sampling had missed the loading spikes that lasted for only a few days.

Research has shown that average annual soil loss can be estimated using only a few site-specific factors, among which is a rainfall-runoff erosivity factor (R). The other factors used to estimate soil loss are soil erodibility, topography, and land use and management. The Universal Soil Loss Equation (USLE) has been revised and is now known as the Revised USLE (RUSLE), based on research by Renard et al. (1991) and Wischmeier and Smith (1978). The rainfall-runoff erosivity factor found in the RUSLE is also used in several nonpoint source models, including the Agricultural Nonpoint Source Pollution Model (AGNPS) (Young et al., 1985). The Water Erosion Prediction Project (WEPP) Hillslope Profile version erosion model is a “new generation” soil erosion model that can be run both as a continuous simulation model and on a single-storm basis. The model requires a large number of data on management practices, which might be difficult to obtain (Singh and Fiorentino, 1996). A procedure derived from the NURP program uses storm frequency and other factors to determine recurrence intervals for instream pollutant concentrations resulting from urban nonpoint source pollution (USEPA, 1984b).

2.4.4 Land Use Data

Landuse data include information on treatments applied to land, current and historical use of the land, spatial and temporal information on land use activities, and changes in land use made before and during a project. Data on these elements are important for evaluating correlations between land surface activities and water quality. Establishing a correlation between a change in water quality and a change in land treatment must be based on both the detection of a water quality trend and detailed information on changes in land use or management, and it requires rigorous statistical analysis (Goodman, 1991; Meals, 1991, 1992). Land treatment can be linked to water quality impacts at the field, subwatershed, watershed, or project level. In general, the larger the drainage area, the harder it is to associate land treatment and water quality. Subwatershed monitoring is the most effective means for demonstrating water quality improvements from a system of BMPs because at this scale the confounding effects of external factors, other polllutant sources, and other BMPs or BMP systems are minimized (Coffey et al., 1993).

Two key points must be considered in nonpoint source monitoring with respect to linking water quality and land treatment. First, weather and season are important confounding influences on nonpoint source activities because they strongly influence the types of land-based activities that can occur, and hence the timing and quantity of runoff from treated lands and the consequential water quality effects. Second, spatial variation must be considered. The location of land treatments relative to surface waters is likely to vary from year to year, and this adds variation to the effect of land treatment on water quality (Meals, 1991).

Correlations between water quality and land treatment can be made much more easily if land use and land treatment monitoring are considered as part of monitoring design in a project's preliminary stages. It is also very important to control for the effects of hydrologic variation. Paired regression is an effective method to control for background variability and is recommended (Meals, 1991, 1992).

Geographic information systems (GIS) are effective management tools for land use data (Meals, 1991). They allow for tracking and manipulating spatial land use data and remarkably improve the ease of visual inspection and comprehension of the data. Data for GIS are available from a variety of sources, including state agencies, GIS user groups, GIS vendors, universities, consultants, conferences, and numerous publications dedicated to GIS topics (Griffin, 1995).

2.4.5 Topographic Data

Topographic data are also required for many nonpoint source monitoring and evaluation efforts, particularly when soil erosion, water runoff, and sedimentation are estimated with models. For example, the USLE includes both slope length and slope steepness factors (Wischmeier and Smith, 1978). AGNPS input includes a slope shape factor, field slope length, channel slope, and channel side slope (Young et al., 1985).

2.4.6 Soil Characteristics Data

Other data such as soil chemistry and soil physical characteristics might be required for some monitoring and evaluation efforts. Recent approaches to assessing the potential for ground water contamination from nonpoint sources have emphasized the need for data such as hydrologic soil group, soil organic carbon content, depth to water, net recharge, aquifer media, and vadose zone characteristics (Aller et al., 1985; Dean et al., 1984).

2.5 PROGRAM DESIGN

Numerous program designs can be used to evaluate the monitoring objectives identified earlier in this chapter. To select the program design, the researcher should develop clear, quantitative monitoring objectives; understand the watershed or waterbody to be monitored; and know something about the locations of and pollutant transport from point and nonpoint sources. In developing the information expectations report described earlier in this chapter, the technical staff will typically decide whether parameter estimation or hypothesis testing is the primary evaluation tool. This choice has an impact on the program design. As an example, balanced designs (e.g., two sets of data with the same number of observations in each set) are generally more desirable for hypothesis testing, whereas parameter estimation might require unbalanced sample allocations to account for spatial and temporal variabilities (Gaugush, 1986). Hypothesis testing is likely to be used in a program evaluation (e.g., water quality before and after pollution controls are implemented), whereas parameter estimation can be applied in assessments when determining pollutant loads from various sources. Hypothesis testing will typically require more intensive databases than those needed for objectives that entail general water quality assessments. As a result, the sampling methodologies required to meet different objectives for the same waterbody may differ considerably.

Most monitoring programs are based on either a probabilistic or a targeted design, or some combination of the two. Probabilistic designs include random selection of station locations and/or sampling events to provide an unbiased assessment of the waterbody. In targeted designs, monitoring sites are selected based on known existing problems or knowledge of upcoming events in the watershed such as installation of a BMP. The most common types of targeted designs employed for the evaluation of nonpoint source pollution sources and BMP systems include monitoring single watersheds, nested watersheds (e.g., above-and-below implementation), two watersheds, paired watersheds, multiple watersheds, and trend stations. Statistical procedures to analyze the data from these study designs are presented in Chapter 4.

Example Objective: *Determine the annual loading of phosphorus from a watershed with no point sources.*

Sampling Methodology: *Assuming no snowmelt inputs and that the majority of phosphorus is delivered under high-flow conditions, the investigator should perform flow-proportional sampling during events. This, of course, assumes that a stage-discharge relationship has been established. Vertical and horizontal concentration and flow profiles should be assessed to determine the need for transect and/or depth-integrated sampling.*

Simply identifying the site location and sampling frequency is not sufficient to describe the where and when of sampling programs. Additional considerations include the depth of sampling, the origins of the aliquot(s) taken in each sample bottle, the time frame over which measurements are made, and others. For example, if a stream is well mixed, a single grab sample from the center of the stream might be sufficient, whereas it might be more appropriate to take an integrated sample from a wider stream. In deeper estuaries, it is a common practice to collect samples near the top and bottom of the waterbody as well as just above and just below the pycnocline. Frequency of sampling should be based on several factors (Sherwani and Moreau, 1975):

- Response time of the system
- Expected variability of the parameter
- Half-life and response time of constituents
- Seasonal fluctuation and random effects
- Representativeness under different conditions of flow
- Short-term pollution events

- Variability and types of the inputs
- Magnitude of response

Examples of sample type classifications include instantaneous and continuous; discrete and composite; surface, soil profile, and bottom; time-integrated, depth-integrated, and flow-integrated; and biological, physical, and chemical. Several existing guidance manuals (Brakensiek et al., 1979; Koterba et al., 1995; Lapham et al., 1995; Platts et al., 1983; Scalf et al., 1981; Shelley, 1979; Shelton, 1994; Shelton and Capel, 1994; USDA-NRCS, 1996; USEPA, 1978b, 1981, 1987a; USGS, 1977) and other reference materials (Wetzel and Likens, 1979) describe these various sample types and the equipment used to collect them.

Selecting an appropriate sampling design for nonpoint source monitoring and evaluation efforts can be a complicated and frustrating experience for the program manager. In addition to balancing multiple (and sometimes competing) objectives, program managers must contend with large variabilities in measured parameters. These variabilities are caused by several factors, including distance to the pollutant source; nonuniform distribution of the pollutant due to physical, biological, or chemical influences; buildup or degradation over time; temporal and spatial variation in background levels; diversity in the biological community; and other nonuniformities such as those in topology, climatic conditions, and waterbody geometry. These factors, in turn, make collecting accurate and unbiased environmental samples more difficult. Biased samples are those which result in consistently higher or lower values than what exists in the waterbody. For example, suspended solids samples taken only during base flow conditions will most likely result in low estimates of annual solids loadings. Accuracy is a measure of how close the sample value is to the true population value. It is necessary to design sampling efforts that meet accuracy requirements while not placing unreasonable burdens on personnel or budgets. Data that are biased or do not meet the project's accuracy requirements are of little use to program managers. An exception might be volunteer data, which often do not meet accuracy requirements but are highly useful in gaining public support for projects.

Other types of sampling uncertainty include random sampling errors and gross errors. Random sampling errors arise from the variability of population units (Gilbert, 1987) and explain why the sample means from two surveys are never equal. Gross mistakes can occur at any point in the process beginning with sample collection and ending with the reporting of study results. Adherence to accepted sampling and laboratory protocols combined with thorough quality control and data screening procedures and experience, dedication, and care will minimize the chances for gross errors.

2.5.1 Probabilistic Designs

In a probabilistic sampling program, the entity about which inferences are made (e.g., watershed) is the population or target population and consists of population units. The sample population is the set of population units that are directly available for measurement. As an example, in a watershed impacted by nonpoint sources, the target population could be defined as storm-event dissolved phosphorus concentrations at the inlets to all impoundments, and phosphorus concentrations in 1-liter grab samples could be population units. Note that both spatial and temporal limits of the water quality variable should be established in defining the target population (Gaugush, 1986). This focuses the sampling program better, in this case eliminating the need to monitor at upstream and in-lake sites, and during baseflow conditions. As a further refinement, the technical staff may define the population units as the dissolved phosphorus concentrations in half-hour composite samples taken during all storms. By sampling and statistically evaluating selected population units, inferences can be made about the entire waterbody.

Simple random sampling

In simple random sampling, each unit of the target population has an equal chance of being selected (Figure 2-3). This type of sampling is appropriate when there are no major trends, cycles, or patterns in the target population (Gilbert, 1987). Random sampling can be applied in a variety of ways, including site selection along the length of a river or areally throughout a lake. Samples may also be taken at a single station using random time intervals. The number of random samples required to achieve a desired margin of error when estimating the mean is (Gilbert, 1987)

$$n = \frac{(t_{1-\alpha/2,n-1} s/d)^2}{1 + (t_{1-\alpha/2,n-1} s/d)^2/N} \tag{2-1}$$

where

n = number of samples,
t = Student's t value,
s = sample standard deviation,
d = absolute margin of error,
N = number of population units, and
α = confidence interval.

If N is large, the above equation can be simplified to

$$n = (t_{1-\alpha/2,n-1} s/d)^2 \tag{2-2}$$

Since the Student's t value is a function of n, both of the above equations are applied iteratively. If the population standard deviation is known, rather than estimated, Equation 2-2 can be further simplified to

$$n = (Z_{1-\alpha/2} \sigma/d)^2 \tag{2-3}$$

where Z is the standard normal deviate and σ is the population standard deviation. In most cases, N is large enough to apply Equation 2-2 or 2-3. Values of Z and t can be found in Appendix D.

Suppose, for example, that the monitoring objective is to estimate the mean dissolved orthophosphate concentration (mg/L as P) during August in a waterbody segment such that there is a 95 percent chance that the mean concentration is within ±0.025 mg/L of the estimated mean. Assuming a population standard deviation of 0.05 mg/L, the number of samples can be estimated using Equation 2-3 as

$$\left(1.96 \times \frac{0.05}{0.025}\right)^2 = 15.4 \approx 16 \textit{ samples}$$

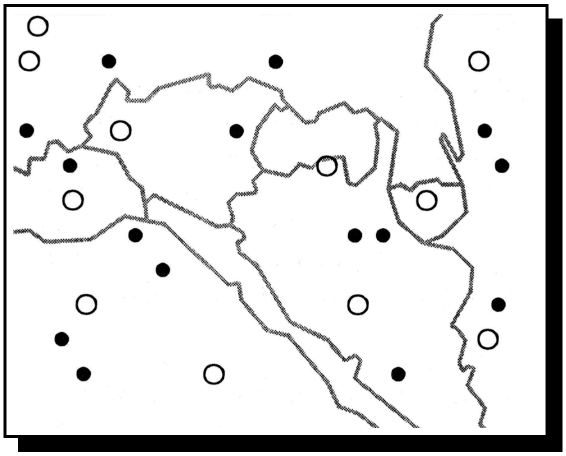


Figure 2-3. Simple random sampling for silviculture. Dots represent harvest sites. All harvest sites of interest are represented on the map, and the sites to be sampled (open dots—○) were selected randomly from all harvest sites on the map. The shaded lines on the map could represent county, watershed, hydrologic, or some other boundary, but they are ignored for the purposes of simple random sampling.

In most cases the standard deviation is not known and Equation 2-2 would be applied. Intuitively, more samples are required due to the uncertainty associated with the standard deviation. To apply Equation 2-2, it is reasonable to initially assume that *n* is equal to some value greater than 16, say 18, which will correspond to a *t* statistic of 2.110. Substituting the above values into Equation 2-2 where the standard deviation now refers to the sample standard deviation yields

$$\left(2.110 \times \frac{0.05}{0.025}\right)^2 = 17.8 \approx 18 \ \textit{samples}$$

Since the computed 18 samples correspond to the initial assumption, no iterations are necessary. In practice, this type of analysis would be performed for several variables and a judgment between sampling size, allowable error, and cost would be made.

Applying any of these equations is difficult when no historical data set exists to quantify the standard deviation. To estimate the population standard deviation, Cochran (1977) recommends four sources:

- Existing information on the same population or a similar population.
- Informed judgment, or an educated guess.
- A two-step sample. Use the first-step sampling results to estimate the needed factors, for best design, of the second step. Use data from both steps to estimate the final precision of the characteristic(s) sampled.
- A “pilot study” on a “convenient” or “meaningful” subsample. Use the results to estimate the needed factors. Here the results of the pilot study generally cannot be used in the calculation of the final precision because the pilot sample often is not representative of the entire population to be sampled.

Gilbert (1987) and Cochran (1977) address additional aspects of simple random sampling. Included in these texts are estimation of the mean and total for sampling with and without replacement, equations for determining the number of samples required for both independent and correlated data, and the impact of measure-ment errors. In most cases, environmental sampling is done without replacement (e.g., aliquots of stream water are not placed back into the stream), *N* is relatively large, the samples are assumed to be independent, and measurement error is ignored, thus making many of these specialized cases less critical. However, the reader should be aware that these issues might become paramount depending on the monitoring objectives and sampling design.

Stratified Random Sampling

In stratified random sampling, the target population is divided into groups called strata for the purpose of obtaining a better estimate of the mean or total for the entire population (Figure 2-4). Simple random sampling is then used within each stratum. Stratification involves the use of categorical variables (e.g., season, flow condition) to group observations into more units that reduce

Example Objective: *Determine the monthly mean total suspended solids concentration (to within ± 15 mg/L at the 95 percent confidence level) for a tributary from an agricultural watershed.*

Sampling Methodology: *Since the concentration may vary with stream depth, width, and flow, the investigator should select a site that is well mixed so that a single grab sample can be taken. If a well-mixed site cannot be found, an integrated sample would be required. Samples would be collected during high and low flow conditions to obtain a representative mean. Random or stratified random samples would then be collected as grab or composite samples depending on the averaging time selected.*

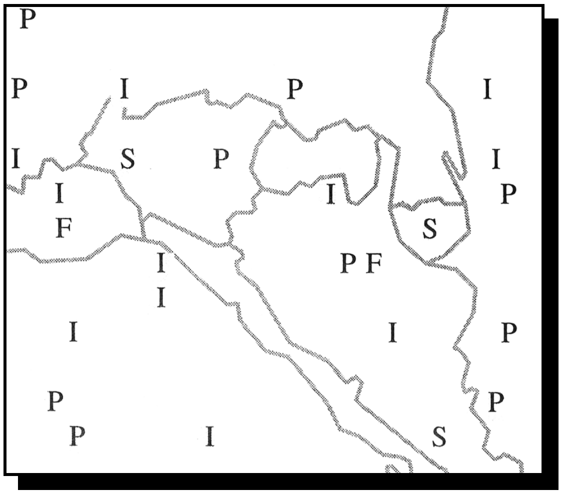


Figure 2-4. Stratified random sampling for silviculture. Letters represent harvest sites, subdivided by type of ownership (P = private nonindustrial, I = industrial, F = federal, S = state). All harvest sites of interest are represented on the map. From all of the sites in one ownership category, sites were randomly selected for sampling (highlighted sites). The process was repeated for each ownership category. The shaded lines on the map could represent county, soil type, or some other boundary, and could have been used as a means for separating the harvest

the variability of observations within each unit. As an example, stratified random sampling can be used to evaluate chemical concentrations in waterbodies when evaluating nonpoint source loadings. One approach would be to stratify stream flow into base and various storm flow periods to account for the energy relationship between precipitation and pollutant generation. Random sampling would then be performed in each stratum.

Cochran (1977) found that stratified random sampling provides a better estimate of the mean for a population with a linear trend, followed in order by systematic sampling (discussed later) and simple random sampling. He also states that stratification normally results in a smaller variance for the estimated mean or total than is given by a comparable simple random sample.
In a stratified random sampling program when N, the number of population units, is large, the optimum number of samples can be estimated with (Cochran, 1977)

$$n = Z^2_{1-\alpha/2} \sum_{h=1}^{L} \frac{W_h s_h^2}{d^2} \qquad (2\text{-}4)$$

where

n = number of samples across all strata,
Z = standard normal variate,
L = number of strata,
W_h = stratum weight,
s_h = sample standard deviation for stratum h,
d = absolute margin of error for weighted mean, and
α = confidence interval.

The stratum weight, W_h, is the relative size of each stratum. Once the total number of samples is determined, the samples may be allocated to each stratum by (Gilbert, 1987)

$$n_h = \frac{n \; W_h s_h}{\sum_{h=1}^{L} W_h s_h} \qquad (2\text{-}5)$$

Alternatively, the samples may be proportionally allocated, with each stratum given a percentage of the total samples in accordance with the stratum size. The above equation allocates more samples to a stratum that is larger or has a higher variability. Cochran (1977) provides an approach for optimizing the sampling when the sampling cost per population unit, c_h, is different among the strata:

$$n_h = \frac{n \; W_h s_h / \sqrt{c_h}}{\sum_{h=1}^{L} W_h s_h / \sqrt{c_h}} \tag{2-6}$$

In general, a larger number of samples would be taken in a stratum that is more variable, larger, or less costly to sample than other strata.

The mean for stratum h, $\overline{x}_h$, is the simple mean of all samples within the stratum. The weighted mean, $\overline{x}_{st}$, is given by

$$\overline{x_{st}} = \sum_{h=1}^{L} W_h \overline{x_h} \tag{2-7}$$

Systematic Sampling

Systematic sampling is used extensively in water quality monitoring programs, usually because it is relatively easy to do from a management perspective. In systematic sampling the first sample is taken from a random starting point (or at a random starting time) and each subsequent sample is taken at a set distance (or time interval) from the first sample (Figure 2-5). For example, if budgetary constraints limit the number of samples to 10 and the objective is to characterize a 10-mile river using systematic sampling, the first observation would be taken randomly in the first river mile. Subsequent samples would be taken at 1-mile increments up the river. In comparison, a stratified random sampling approach would divide the river into 10 1-mile segments (strata) and one random sample would be taken in each segment.

Gilbert (1987) recommends systematic sampling when estimating long-term trends, defining seasonal or other cycles, or forecasting pollution concentrations. In general, systematic sampling is superior to stratified random sampling with one or two samples per stratum for estimating the mean (Cochran, 1977). Gilbert (1987) reports that systematic sampling is equivalent to simple random sampling in estimating the mean if the target population has no trends, strata, or correlations among the population units. Estimates of variance from systematic samples may differ from those determined from random samples. Cochran (1977) notes that “on the average the two variances are equal.” However, Cochran also states that for any single population for which the number of sampling units is small, the variance from systematic sampling is erratic and may be smaller or larger than the variance from simple random sampling.

Gilbert (1987) cautions that any periodic variation in the target population should be known before establishing a systematic sampling program.

Sampling intervals equal to or multiples of the target population's cycle of variation may result in biased estimates of the population mean. Systematic sampling can be designed to capitalize on a periodic structure if that structure can be characterized sufficiently (Cochran, 1977). A simple or stratified random sample is recommended, however, in cases where the periodic structure is not well known or where the randomly selected starting point is likely to have an impact on the results (Cochran, 1977). Quantitative procedures for estimating the population mean and variance from systematic sampling data are presented by Gilbert (1987).

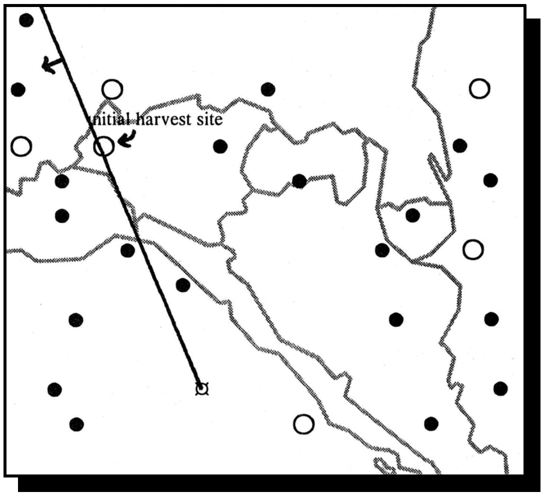


Figure 2-5. Systematic sampling for silviculture. Dots (● and ○) represent harvest sites of interest. A single point on the map (¤) and one of the harvest sites were randomly selected. A line was stretched outward from the point to (and beyond) the selected harvest site. The line was then rotated about the map and every fifth dot that it touched was selected for sampling (open dots—○). The direction of rotation was determined prior to selection of the point of the line's origin and the beginning harvest site. The shaded lines on the map could represent county

Gilbert (1987) notes that assumptions about the population are required in estimating population variance from a single systematic sample of a given size. However, there are systematic sampling approaches that do support unbiased estimation of population variance, including multiple systematic sampling, systematic stratified sampling, and two-stage sampling (Gilbert, 1987). In multiple systematic sampling more than one systematic sample is taken from the target population. Systematic stratified sampling involves the collection of two or more systematic samples within each stratum.

Cluster Sampling

Cluster sampling is applied in cases where it is more practical to measure randomly selected groups of individual units than to measure randomly selected individual units (Gilbert, 1987). In cluster sampling, the total population is divided into a number of relatively small subdivisions, or clusters, and then some of these subdivisions are randomly selected for sampling (Figure 2-6). For one-stage cluster sampling, the selected clusters are sampled totally. In two-stage cluster sampling, random sampling is then performed within each cluster (Gaugush, 1986). An example of one-stage cluster sampling is the collection of all macroinvertebrates on randomly selected rocks within a specified sampling area. The stream bottom might contain hundreds of rocks with thousands of organisms attached to them, thus making it difficult to sample the organisms as individual units. However, it is often possible to randomly select rocks and then inspect every organism on each selected rock.

Gaugush (1986) states that the "analysis of cluster samples requires the estimation of variance at two levels, the between-cluster variability and the within-cluster variability. The total variability is a recombination of these two levels." Freund (1973) notes that estimates based on cluster sampling are generally not as good as those based on simple random samples, but they are more cost-effective. As a result, Gaugush believes that the difficulty associated with analyzing cluster samples is compensated for by the reduced sampling requirements and cost. Cochran (1977) discusses one-stage cluster sampling for clusters of either equal or unequal sizes and provides equations for determining the optimal population unit size using the relative sizes of possible population units, the variance among the population unit totals, and the relative cost of measuring one population unit. He notes that many factors come into play when determining optimal population size, including cost versus unit size.

Two-stage Sampling

Two-stage sampling involves dividing the target population into primary units, randomly selecting a subset of these primary units, and then taking random samples (second-stage units) within each of the selected primary units. This is a common practice when a large sample is taken and then a smaller aliquot is actually measured from the original sample. The process of subsampling introduces additional uncertainty and becomes significant if the pollutant is in particulate form and very small subsamples are used (Gilbert, 1987).

Two-stage sampling might also include systematic sampling within a randomly selected subset of the population primary units. For example, if the target population is the average annual pollutant concentration in a stream, the primary units could be daily average concentrations (n = 365). A subset of these daily concentrations (e.g., n = 24) could be selected at random for further systematic sampling of hourly concentrations. For example if four systematic, hourly samples could be taken on each of 24 different days, with the hour for the first sample determined randomly, followed by three more hourly samples taken every sixth hour, 96 hourly composite samples would be available for the calculation of the population mean and variance.

Cochran (1977) describes two-stage sampling in great detail and presents methods for determining the mean and variance in two-stage sampling with units of equal size. In Cochran's discussion, he notes that if all population units are sampled, the formula for estimating the variance is the same as that used to estimate the variance for proportional stratified random sampling. This means that two-stage sampling is a type of incomplete stratification, with the primary units treated as strata.

For further information regarding two-stage (and three-stage) sampling, the reader is referred to Gilbert (1987) and Cochran (1977). The authors provide equations for estimating the number of samples (primary units) and subsamples for two conditions: (1) primary units of equal size and (2) primary units of unequal size. Equations for estimating the mean and total values in composite samples of equal- and unequal-sized units are also provided. The authors also provide equations for calculating the number of composites and composite subsamples needed.

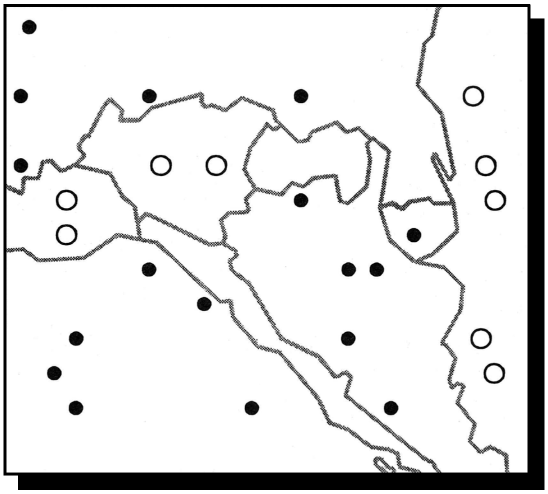


Figure 2-6. Cluster sampling for silviculture. All harvest sites in the area of interest are represented on the map (closed {●} and open {○} dots). The shaded lines on the map represent county boundaries. Some of the counties were randomly selected, and all harvest sites within those counties (open dots - ○) were selected for sampling. Some other type of boundary, such as soil type or watershed, could have been used to separate the harvest sites for the sampling process.

Double Sampling

Double sampling is often used when two techniques exist for measuring a pollutant. Initially, both methods are used. Then, after a correlation has been established, only the cheaper or simpler technique is used. Gilbert (1987) provides an approach for calculating the sample size when the cost and variability associated with both methods has been determined during the initial sampling. This same procedure can also be used when it is less expensive to measure a surrogate variable (Gilbert, 1987). This technique can be used for stratification, ratio estimates, and regression estimates (Cochran, 1977).

Regression analyses are used to predict values for one variable (i.e., the dependent variable) using one or more independent variables based on a mathematical relationship. As an example, total suspended solids concentration is typically a covariate of total phosphorus concentration in watersheds impacted by agricultural runoff. Measurement of total suspended solids may help increase the precision of total phosphorus estimates. Gaugush (1986) discusses sampling to support regression analyses using spatial or temporal gradients as the independent variable, the latter being for trends over time. Some key points in his discussion related to using a spatial independent variable are as follows:

- Whenever the type of relationship (e.g., linear, log-linear) is known, relatively few sampling points are needed along the gradient. More samples may then be used as replicates.

- Whenever the relationship is not known, more sampling points are needed along the gradient. More replicates are also needed to test the proposed model.

- It is usually acceptable to place sampling points equal distances from each other along the gradient as long as the sampling does not fall in step with some natural phenomenon that would bias the data collected.

Some key points in the discussion regarding time sampling are as follows:

- Time can be used either as a covariate or as a grouping variable. Grouping by time might be desirable when changes in the variable of interest either are small over time or occur only during short periods with long periods of little or no change.

- Considerations in using time as a covariate are similar to those for spatial gradients, but (1) time is usually only a surrogate for other variables that truly affect the variable of interest, and (2) the relationship with time is likely to be complex.

- If time is to be used as a covariate, relatively frequent sampling will be needed, with some replication within sampling periods. Random sampling within the periods is also recommended.

The sampling designs most common to environmental monitoring are summarized in Table 2-2.

2.5.2 Targeted Site Location Study Designs

Paired and nested paired watershed approaches are the two most appropriate approaches when trying to evaluate the impact or benefit of a BMP or system of BMPs at the watershed scale (Spooner et al., 1985). A nested paired watershed design (Figure 2-7A) is sometimes referred to as an "above-and-below" design where one monitoring station is located above the treatment area and one station is located below the treatment area. The paired watershed design (Figure 2-7B) is based on identifying two watersheds where one watershed is the control and the second is the treatment. In both study designs, data are collected before treatment (calibration) and after treatment is implemented so that differences between watersheds (or nested watersheds) can be evaluated. The key advantage of these two approaches is that the variation due to

Table 2-2. Applications of six sampling designs to estimate means and totals.

Sampling Design	Conditions for Application
Simple Random Sampling	Population does not contain major trends, cycles, or patterns of contamination.
Stratified Random Sampling	Useful when a heterogeneous population can be broken down into parts that are internally homogeneous.
Two-stage Sampling	Needed when measurements are made on subsamples or aliquots of the field sample.
Cluster Sampling	Useful when population units cluster together and every unit in each randomly selected cluster can be measured.
Systematic Sampling	Usually the method of choice when estimating trends or patterns of contamination over space. Also useful for estimating the mean when trends and patterns in concentrations are not present.
Double Sampling	Useful when there is a strong linear relationship between the variable of interest and a less expensive or more easily measured variable.

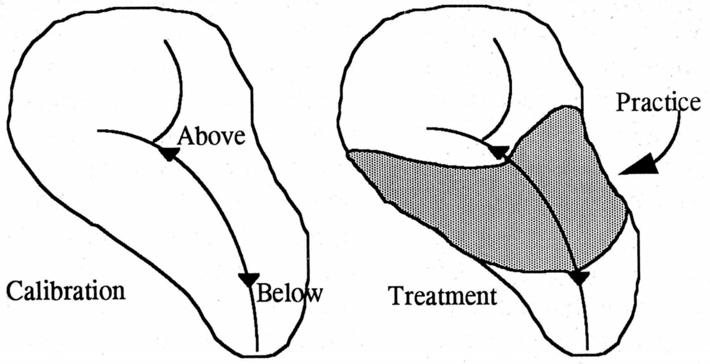
Source: After Gilbert, 1987.

year-to-year climatic differences and differences between watersheds are statistically controlled, provided that a sufficient calibration period has been used. Clausen (1991) states that the cost of conducting a paired watershed experiment in Vermont ranged from $30,000 to $50,000 per year for 3 or 4 years. This cost included continuous discharge and water sampling, as well as the analysis of approximately six water quality characteristics.

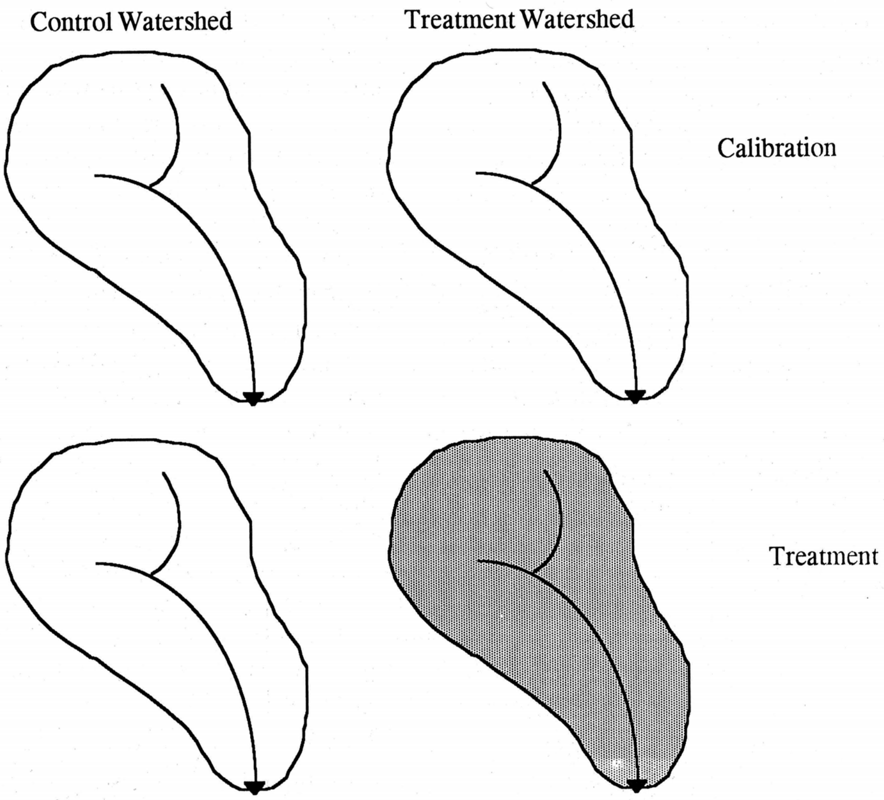
In St. Albans Bay, Vermont, in another RCWP, two small watersheds received proper manure management during a 2-year calibration period, followed by a period in which one watershed received winter-spread manure (Clausen, 1985). This is an interesting example of the paired watershed approach since BMPs were removed from, instead of applied to, a watershed after the calibration period. Data from this type of nested paired or paired watershed design can be evaluated by an analysis of covariance as described by USEPA (1993c). Unfortunately, both study designs are limited because the experiment is not repeated to account for spatial variability, and transferability of BMP effectiveness to other regional watersheds is not appropriate (MacDonald et al., 1991).

Nested watershed designs can also be used to document the severity of a nonpoint source pollution problem. In an example from the Rock Creek, Idaho, RCWP, paired data were collected using an upstream-downstream approach. These data were used in regressions of water quality against time.

The downstream concentrations (below the nonpoint pollution source) were adjusted for upstream concentrations (above the nonpoint pollution source), transformed, and then regressed against time as a continuous variable (Spooner et al., 1986). Results of this approach indicated that decreasing pollutant concentrations from nonpoint pollution sources were due to implementation of BMPs.



A) Nested Paired Watershed Study Design



B) Paired Watershed Study Design

Figure 2-7. Nested paired and paired watershed study designs.

Single-watershed designs, which collect data before and after BMP implementation, and two-watershed designs, which collect data after BMP implementation in one watershed, should generally be avoided for evaluating BMP effectiveness. The single-watershed design does not account for year-to-year climatic variability. The two-watershed design does not account for differences between watersheds since no calibration data are collected.

An alternative approach, when collecting data during a calibration period is not viable, is to use a multiple-watershed design, in which numerous watersheds are monitored. In this design, multiple watersheds in a region are selected, including some that have a particular BMP implemented and others that do not have the BMP implemented. Alternatively, numerous paired upstream and downstream stations (i.e., nested watersheds) are selected. In the case of paired upstream and downstream stations, the designation of controls or treatments is not random, and it is necessary to add additional station pairs where no treatment or BMP is implemented (MacDonald et al., 1991). By monitoring numerous watersheds, the true variability between watersheds is considered and the results from this study design can be transferred to other watersheds in the region. Fifteen paired stations were established in the Snohomish River basin (Washington State) to determine the effect of commercial agriculture on water quality along with other objectives over a 3-year period (Luchetti et al., 1987). The pairs varied considerably in terms of stream size and agricultural activity. Combining the monitoring data with land use and BMP implementation data, the project documented the impact of commercial agriculture on water quality.

Use of trend stations, or long-term ambient monitoring, is based on establishing monitoring stations that are routinely monitored. This type of study design is generally most appropriate for watersheds where a variety of BMPs are being implemented over a period of time or gradual water quality changes are expected. The difficulty in using trend stations is developing a causal link between water quality and the various land use activities. To use trend stations, variables associated with land treatment, hydrology, and meteorology should be accounted for to increase the likelihood of successful documentation of water quality-BMP relationships. The long-term commitment required from management to monitor these stations is one of the key disadvantages of this approach. The U.S. Geological Survey has systematically sampled the national stream quality accounting network (NASQAN) once a month for more than 20 years to monitor the water quantity and quality (Smith et al., 1987).

One key to establishing the study design, which is often overlooked, is site selection. Site location and establishment are discussed in several existing monitoring guides and texts (Brakensiek et al., 1979; Ponce, 1980a; USEPA, 1978b, 1981; USGS, 1977; Wetzel and Likens, 1979). Few differences exist between nonpoint source site location strategies and the approaches discussed in these documents. Within any given budget, site location is a function of water resource type, monitoring objectives, and data analysis plans. When evaluating the effectiveness of nonpoint source control measures, it might be necessary to locate monitoring sites above known point sources to remove them as confounding influences in the study. Additional considerations in site selection are site accessibility and landowner cooperation in data collection efforts (e.g., farm management records). It is strongly recommended that nonpoint source monitoring stations be located near or at USGS gaging stations, when possible, due to the extreme importance of obtaining accurate flow records for estimating pollutant loads. In the absence of a USGS gaging station, monitoring stations should be located at sites that offer adequate flow monitoring capabilities. Some station requirements may be such that, with careful station siting, one particular station can meet multiple monitoring objectives. Caution should be exercised, however, to avoid compromising the worth of a station for the sake of false economy.

For evaluating the overall background or performing a problem assessment, a panel of federal and state monitoring professionals (USEPA, 1975) determining several points for establishing site locations for physical and chemical water column sampling, which should be considered as appropriate. The process of site selection for biological monitoring is described in Chapter 3.

- Sites should be located at representative sites in mainstem rivers, estuaries, coastal areas, lakes, and impoundments. These sites can be used to characterize the overall quality of the area's surface waters and will provide water quality baselines against which progress can be measured.

- Sites should be located in water quality-limited and major water use areas. Sites in water quality-limited areas can be used to evaluate the overall pollution control strategy and BMP system effectiveness. Sites in major water use areas, such as public water supply intakes, commercial fishing areas, and recreational areas, serve a dual purpose—public health protection and overall water quality characterization.

- Sites should be located upstream and downstream from representative land use areas (e.g., mining, silviculture) and morphologic zones. These sites can be used to compare the relative effects of pollution sources and morphologic zones on water quality and to document baseline water quality.

- Sites should be located at the mouths of major or significant tributaries to mainstem streams, lakes, impoundments, estuaries, or coastal areas. Data from these sites, when taken in concert with permit monitoring data and intensive survey data, can be used to determine the major sources of pollutants to the area's major waterbodies. By comparison with other tributary data, the relative magnitude of the pollution sources can be evaluated and problem areas can be identified.

- Sites should be located to measure the input and output of nutrients and other pertinent substances into and from waterbodies (i.e., lakes, impoundments, estuaries, or coastal areas) that exhibit eutrophic characteristics, as well as at critical locations within the waterbody. The information from these stations, when taken in combination with the pollution source data, can be used to establish cause-and-effect relationships, identify problem areas, and indicate appropriate corrective measures.

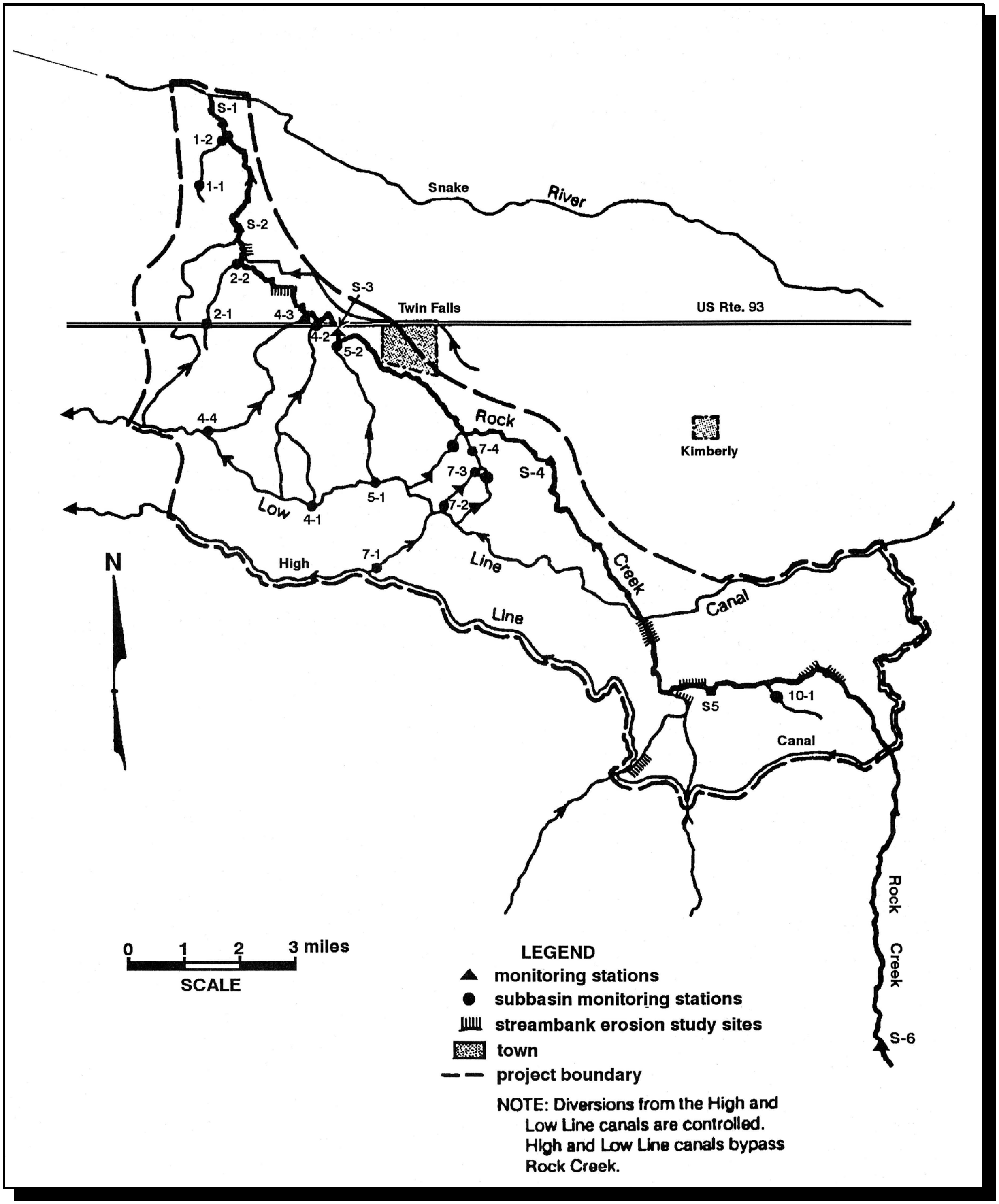
Sediment sampling sites should be located in sink areas as determined by intensive surveys, reconnaissance surveys, and historical data. A major concern of sediment monitoring is to assess the accumulation of toxic substances and sediment-bound nutrients. The location for a sediment sampling site should be chosen by considering the sediment mechanics and the hydrological characteristics of the waterbody (USEPA, 1975).

2.6 EXAMPLE PROGRAM DESIGN

The RCWP includes several examples of nonpoint source monitoring and evaluation strategies. Two project strategies are described here. Several additional examples are provided in Appendix C.

The Idaho RCWP's major focus was to control sediment from irrigation return flows. Using a targeted study design, seven ambient monitoring stations (Figure 2-8) were used (Clark, 1986):

S-1: Near mouth - integrated all pollution sources flowing into Rock Creek and measured the pollutant load that going into the Snake River (river mile (RM) 0.75). Water quality, benthic macroinvertebrates, and fisheries data were collected.



NOTE: Diversions from the High and Low Line canals are controlled. High and Low Line canals bypass Rock Creek.

Figure 2-8. Map of the Rock Creek Rural Clean Water Program study area, Twin Falls County, Idaho. (Source: Clark, 1986)

S-2: At Poleline Road - a benthic invertebrate and fisheries monitoring site as well as water quality (RM 3.75).

S-3: Above Highway 93 - below the confluence of the high-priority agricultural drains and city of Twin Falls urban runoff (RM 7.3). Water quality, benthic macroinvertebrates, and fisheries data were collected.

S-4: At Twelvemile - above the influence of Twin Falls urban area and the high-priority drains (RM 13.5). Water quality, benthic macroinvertebrates, and fisheries data were collected.

S-5: At 3500 East Road - a benthic invertebrate and fisheries monitoring site only (RM 21.1).

S-6: Near Rock Creek townsite - measured the quality of the natural surface water above the irrigation tract (RM 30.3). Water quality, benthic macroinvertebrates, and fisheries data were collected.

C-1: Twin Falls Main Canal - source of water for the irrigation tract. Only water quality data were collected.

Intensive monitoring stations were placed on irrigation drains to track changes in sediment load and associated pollutants close to their source and associated BMPs. In this way, changes in water quality due to the RCWP could be detected. Nineteen stations were located in six subbasins (Figure 2-8). Stations measured the source of water to the subbasins (7-1, 5-1, 4-1, 4-3, 2-1, and 1-1), the input of the subbasins to Rock Creek (7-7, 7-4, 5-2, 4-2, 4-3, 2-2, and 1-2), and key intermediate sites (7-2, 7-3, and 7-6). Additional stations were added in other subbasins as they were needed (2-3, 2-4, and 10-1).

The St. Albans Bay, Vermont, RCWP project used a four-level monitoring and evaluation program to meet three objectives (Vermont RCWP Coordinating Committee, 1986):

- Document changes in the water quality of specific tributaries within the watershed resulting from implementation of manure management practices.

- Measure changes in suspended sediment and nutrients entering St. Albans Bay resulting from implementation of water quality management programs within the watershed.

- Evaluate trends in the water quality of St. Albans Bay and the surface waters within the St. Albans Bay watershed during the period of the St. Albans Bay RCWP Watershed Project. Monitoring sites for all four levels of monitoring and evaluation are shown in Figure 2-9. The Level 1 bay sampling was designed to determine long-term water quality trends in St. Albans Bay over the life of the project (Vermont RCWP Coordinating Committee, 1984). The Level 2 tributary sampling was designed to determine the long-term water quality trends for the major tributaries including the Bay and the St. Albans City wastewater treatment plant (Vermont RCWP Coordinating Committee, 1984). The Level 3 monitoring was directed toward evaluating the effect of best manure management practices on the quality of surface runoff from individual fields; Level 4 was designed to supplement the Level 2 monitoring by sampling additional tributaries to St. Albans Bay and to isolate subunits within the Level 2 subwatersheds (Vermont RCWP Coordin-ating Committee, 1984).

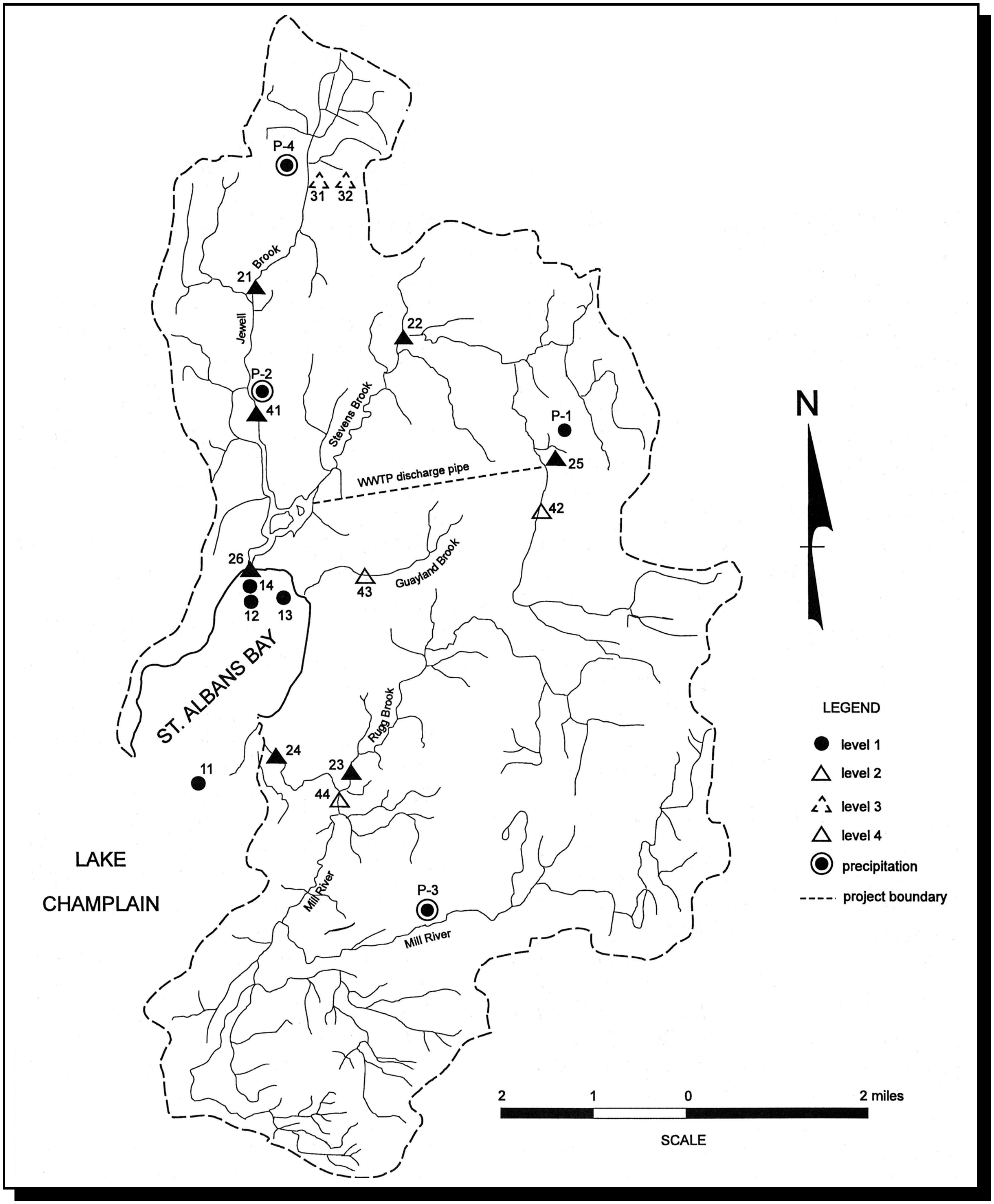


Figure 2-9. St. Albans Bay watershed, Franklin County, Vermont, sampling locations. (Source: Vermont RCWP Coordinating Committee, 1986)

2.7 ROLES AND RESPONSIBILITIES

Designing and implementing a monitoring program is an interdisciplinary and interagency activity. In many cases, technical staff will need to integrate "new" monitoring with what is already being done in order to demonstrate to program managers that duplicate work is not proposed. The most effective way to achieve this goal is to bring all the involved agencies and other stakeholders in the monitoring effort together. One or a few agencies acting as project coordinator(s) should seek to obtain an agreement from each involved party with respect to their role(s) and responsibilities in the performance of the project. These agreements can be formalized as commitments and specified in the quality assurance project plan, which is discussed at greater length in Chapter 5.

Such coordinated cooperation permits each involved party to offer the results of its ongoing activities to the monitoring effort and lessens the burden on the proposed budget. For example, the U.S. Geological Survey might already have a gaging station in place and the Natural Resources Conservation Service might already have a tracking system for BMPs in place. Other agencies, including the U.S. Fish and Wildlife Service and EPA, might have other ongoing monitoring programs. When multiple agencies are involved in the monitoring program, each can benefit from the efforts of the others.

2.8 QUALITY ASSURANCE PROJECT PLANNING

An integral part of the design phase of any nonpoint source pollution monitoring project is the development of a quality assurance project plan (QAPP). The QAPP is a critical document for the data collection effort inasmuch as it integrates the technical and quality aspects of the planning, implementation, and assessment phases of the project. The QAPP documents how quality assurance (QA) and quality control (QC) elements will be implemented throughout the life of a project. It contains statements about the expectations and requirements of those for whom the data are being collected (i.e., the decision makers) and provides details on project-specific data collection and data management procedures that are designed to ensure that these requirements are met. Development and implementation of a QA/QC program, including preparation of a QAPP, can require up to 10 to 20 percent of project resources (Cross-Smiecinski and Stetzenback, 1994), but this cost is recaptured in lower overall costs due to the project's being well planned and executed. A thorough discussion of QA/QC is provided in Chapter 5.

2.9 CHEMICAL AND PHYSICAL MONITORING

Chemical and physical monitoring and the mechanics of sampling are important topics and need to be considered as carefully as other monitoring topics discussed in this guide, such as data analysis and biological monitoring. However, these aspects of monitoring are covered in detail in other documents (e.g., USDA-NRCS, 1996; USGS, 1977) and it would be redundant to duplicate the information here. Therefore, these types of monitoring are only briefly mentioned here.

Important topics related to chemical and physical monitoring and sampling procedures that managers of nonpoint source pollution monitoring programs should consider include the following:

- **Type of sample**. Water quality varies temporally and spatially, and samples must be taken that will accurately reflect overall water quality and overall water quality impacts of nonpoint source pollutants. There are four basic types of samples to consider—grab, composite, integrated, and continuous (USDA-NRCS, 1996):

  Typically, a grab sample is a sample taken at one place a single time. Care should be taken to make sure that a grab sample is representative. If there is spatial variability (e.g., across a stream, at different depths in a lake) or temporal variability (e.g., during a storm event) it might be more appropriate to take a composite or time-integrated sample rather than a grab sample.

Composite samples consist of a series of grab samples, usually collected in the same location but at different times with the results averaged. Composite samples are usually either time-weighted or flow-weighted. Time-weighting means that a fixed volume is collected at a predetermined time interval. Flow-weighting means that a sample is taken after a specified quantity of water has passed the monitoring station. Both types of composite sampling are amenable to automatic sampling equipment. Composite samples are appropriate for most monitoring objectives.

Integrated samples account for variations in water quality with depth or distance from a stream bank at a monitoring station. Subsamples are taken at various depths or distances from the stream bank, and integrated into a single sample.

Continuous sampling requires electronic measuring devices and is therefore limited to variables that are amenable to this type of sampling, such as dissolved oxygen, conductivity, pH, and salinity. It is generally not suitable for measurements of metals, organics, or pesticides. Continuous sampling is typically used for research and fate and transport studies.

Some factors that influence the type of sample to collect include the objectives of the study, waterbody type, and variables to be sampled.

- **Type of sample collection**. Samples can be collected manually or with automated equipment. Sampling location, sample site accessibility, and staffing are factors to consider when determining which approach to use.

- **Type of sample collection equipment**. Sampling equipment can be either mechanically operated or powered, and the use of one or the other approach again depends on project-specific considerations and constraints. Commonly used sampling equipment includes flow recorders, staff gauges, and precipitation gauges.

- **Station type**. Various monitoring stations might be necessary to measure the variables of interest. Discharge stations might be installed to measure runoff from a sampling plot in a field or at the edge of a field, or to measure stream discharge. Other monitoring stations might be necessary to collect water samples, record precipitation, analyze soil water, assess biological factors, or monitor sediment.

- **Sampling equipment operation and maintenance**. It is important to ensure that all sampling equipment is in good operational condition prior to sampling and during sampling to ensure that reliable data are being collected. The use of automated sampling equipment does not mean that project staff are relieved of the responsibility to regularly check equipment operation. Staff should be thoroughly trained to use and maintain sampling equipment properly.

- **Record keeping**. Proper record keeping is important to make the process of data analysis less burdensome and to aid in tracking any anomalies in data to possible influences, such as equipment malfunctions or variations in sample collection timing. Detailed records are also valuable when writing reports and preparing presentations.

2.10 RECOMMENDED REFERENCES

Important monitoring references that should be consulted include the following:

American Public Health Administration. 1995. *Standard methods for the examination of water* *and wastewater*. 19th ed. American Public Health Association, Washington, DC.

Discussion of how to collect samples and the required volume of sample material for numerous water quality parameters.

Bauer, S.B., and T.A. Burton. 1993. *Monitoring protocols to evaluate water quality effects of grazing management of western rangeland streams*. EPA 910/R-93-017. Submitted to U.S. Environmental Protection Agency, Region 10, Water Division, Surface Water Branch, by Idaho Water Resources Research Institute, University of Idaho, Moscow, ID. October.

Temperature, nutrients, bacteria, stream channel morphology, stream bank stability, sediment, streamside vegetation. For each, parameters to measure, sample collection procedures, sample analysis.

Clark, W.H. 1990. *Coordinated nonpoint source water quality monitoring program for Idaho*. Idaho Department of Health and Welfare, Division of Environmental Quality, Boise, ID. January.

Appendix with suggested monitoring parameters and protocols, including suggested protocols for various types of BMP implementation and pollutant sources and transport mechanisms.

MacDonald, L.H., A.W. Smart, and R.C. Wissmar. 1991. *Monitoring guidelines to evaluate effects of forestry activities on streams in the Pacific Northwest and Alaska*. EPA/910/9-91-001. U.S. Environmental Protection Agency, Region 10, Seattle, WA.

Forestry focus; parameter selection; discussion of many parameters, including a definition, relation to designated uses, how the parameter responds to management activities, parameter-specific measurement notes, applicable standards, present uses of the parameter, and parameter assessment. Parameter recommendations for various land treatments.

USDA. 1979. *Field manual for research in agricultural hydrology*. Agricultural Handbook 224. U.S. Department of Agriculture, Washington, DC.

USDA-NRCS. 1996. *Water quality monitoring*. U.S. Department of Agriculture, Natural Resources Conservation Service, Washington, DC.

Numerous recommendations for good references on a variety of sampling topics. Also includes tables with recommendations of variables to measure based on the above considerations. Topics covered include variable selection, sample types (grab, composite, integrated, continuous), station type (discharge, concentration, precipitation, soil water, biotic, sediment), sample collection (volume), sample preservation.

USGS. 1977. *National handbook of recommended methods for water-data acquisition*. U.S. Geological Service, Office of Water Data Coordination, Reston, Virginia.