Tech NOTES 2

National Nonpoint Source Monitoring Program

July 2005

Steven A. Dressing and Donald W. Meals. 2005. Designing water quality monitoring programs for watershed projects, Tech Notes 2, July 2005. Developed for U.S. Environmental Protection Agency by Tetra Tech, Inc., Fairfax, VA, 20 p. Available online at *https://www.epa.gov/polluted-runoff-nonpoint-source-pollution/nonpoint-source-monitoring-technical-notes*.

Through the National Nonpoint Source Monitoring Program **(NNPSMP)**, states monitor and evaluate a subset of watershed projects funded by the Clean Water Act Section 319 Nonpoint Source Control Program.

The program has two major objectives:

1. To scientifically evaluate the effectiveness of watershed technologies designed to control nonpoint source pollution
2. To improve our understanding of nonpoint source pollution

NNPSMP Tech Notes is a series of publications that shares this unique research and monitoring effort. It offers guidance on data collection, implementation of pollution control technologies, and monitoring design, as well as case studies that illustrate principles in action.

Designing Water Quality Monitoring Programs for Watershed Projects

Introduction

This technical note describes an approach to designing water quality monitoring programs for watershed projects. The primary focus is on the design of monitoring to evaluate the effects of watershed land treatment programs on water quality.

Monitoring is often defined 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 (ITFM, 1995). Monitoring program design should be an integral part of watershed project planning from the very beginning.

Before a monitoring program is designed and implemented, goals should be defined clearly and available information from past and current monitoring efforts should be critically reviewed (MacDonald, Smart, and Wissman, 1991; USDA, 1996; USEPA, 1997a). Constraints such as funding levels, personnel availability, site accessibility, and sample collection logistics must be factored into the initial design of the monitoring program to ensure that it is feasible. Long-term commitments of time and resources, though difficult to nail down, are essential to the consistent and uninterrupted collection of monitoring data.

Define Goals

Monitoring is carried out to support watershed projects for a number of reasons, including the following (USEPA, 1997a):

- To identify water quality problems, use impairments, causes, and pollutant sources
- To assess permit compliance
- To assist program development and management

- To respond to emergencies
- To validate or calibrate models
- To conduct research
- To develop TMDLs and load/wasteload allocation
- To assess use support status
- To track trends
- To track management measure implementation
- To assess the effectiveness of watershed projects

The need for such monitoring will depend on the goals of the watershed project, the complexity of the water quality issues to be addressed, the size and land uses within the project area, and the current availability of watershed data. For example, although it might be known that eutrophication is a problem in a particular lake, the development of an effective management measure implementation plan could require additional monitoring to fully assess the impacts of the eutrophication, to quantify nutrient sources, and to identify the pathways of nutrient delivery to the lake and nutrient cycling within the lake.

Monitoring should be designed to meet the main goals of the project. Because the focus of this technical note is on evaluation of a watershed project through monitoring, the most relevant monitoring goals are to track management measure implementation, look for trends, and assess watershed project effectiveness. Land treatment monitoring goals might include the following:

- To find out when and where management measures are implemented and operational
- To determine whether management measures are working as planned
- To determine the degree of pollution control achieved by the management measures
- To measure the pollutant contributions from areas where management measures are not implemented
- To discover unplanned activities that could affect project success

Trend analysis and watershed effectiveness monitoring goals might include the following:

- To document pre-implementation water quality conditions
- To measure changes in water quality due to implementation of management measures
- To develop information to guide changes in the implementation plan if water quality goals are not achieved
- To measure the pollutant removal efficiencies of specific management measures
- To measure water quality changes in subwatersheds

- To document changes in pollutant load at the watershed outlet

Where multiple goals are established for a monitoring program (e.g., load allocation and project effectiveness), specific design details must be considered carefully to ensure that all goals are suitably addressed. For example, instrumented sampling locations should be selected to maximize the potential for addressing more than one goal cost-effectively.

Review Available Information and Monitoring Efforts

A good watershed monitoring program must be based on a thorough understanding of the system(s) being monitored. Collecting and evaluating all available information and data from other monitoring efforts lays an important foundation for such an understanding. Currently available information should be reviewed before new data are collected to assess its potential use in characterizing the watershed and achieving monitoring goals. Exploratory analysis of existing data might yield information that can help locate hot spots or critical areas, identify important covariates, or account for such characteristics as seasonality in the design of the monitoring program (USEPA, 2005a).

Knowledge of the variability (i.e., "noise") of the systems being monitored is very important because variability has a profound effect on the design and cost of the monitoring program and on the ability of the program to detect change reliably. Monitoring sites for which long periods of record exist can be particularly useful in determining the sampling frequencies needed to make inferences with confidence using appropriate statistical analysis.

Existing data can also be helpful in designing the management measure implementation plan. For example, stream data for relatively homogenous watersheds might be helpful in assessing pollutant delivery coefficients for current land uses. Comparing these coefficients with literature values might provide a crude indication of the extent to which management improvements could reduce pollutant delivery.

When reviewing historical data, it is important to explore any relationships that might exist between water quality data and land use or land management data. Abrupt changes in water quality parameter values could be related to the addition of or improvement to a point source discharge such as a wastewater treatment plant, changes in impervious surface percentage, increases or decreases in livestock herd sizes, changes in agricultural crop production or livestock types, hydromodification or bridge construction, urbanization, or other land use or land management changes.

Design the Monitoring Program

The specifics of a successful monitoring program depend on the monitoring goals, the availability of existing data and monitoring efforts, the time frame for yielding results, the variability of the system monitored, the types of variables tracked, funding, and the priorities of program managers. Just as with automobiles there are basic requirements (e.g., engine, chassis, wheels, seats, brakes) and options (e.g., power seats, anti-lock brakes, CD player, aluminum wheels), a monitoring program has both basic requirements and options. The basic requirement is that the monitoring goals must be met with a comfortable degree of confidence, while the options (e.g., additional sampling locations, parameters, and samples; remote access where not necessary; the latest laboratory equipment) are seemingly endless. A no-frills monitoring program that meets established goals should be the base, to which enhanced capabilities are added as resources and management allow. Even with generous budgets, no monitoring should be conducted without carefully considering the use of the data to be collected.

Statistical Design

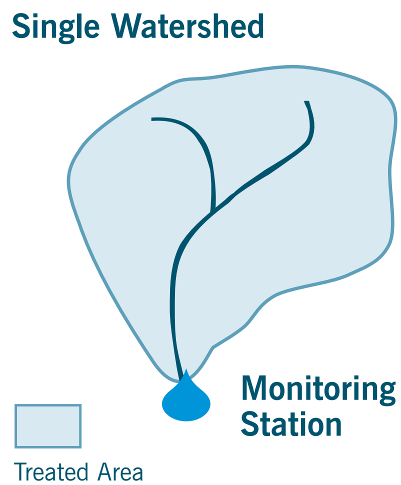
The statistical design must be chosen before other monitoring details (e.g., scale, number of sampling sites, monitoring station type, sampling frequency) can be determined (USDA, 1996). Gilbert (1987) describes four basic options for sampling: haphazard sampling, judgment sampling, search sampling, and probability or statistical sampling. Haphazard and judgment sampling should be avoided; search sampling can be used to characterize a watershed and locate major pollutant sources; and probability sampling should be used to evaluate watershed projects. Types of probability sampling include simple random sampling, stratified random sampling, two-stage sampling, cluster sampling, systematic sampling, and double sampling (Gilbert, 1987; USEPA, 1997a). Simple random sampling is useful for estimating means and totals for homogenous populations, while stratified random sampling is recommended to estimate means and totals if the data are stratified into distinct levels or groupings (e.g., seasonal). Systematic sampling with the starting date selected randomly is recommended over random sampling when the goal is to track a variable over time or space.

When evaluating the effectiveness of watershed projects, the emphasis should be on testing a hypothesis rather than estimating parameters. For example, the null hypothesis might be that phosphorus loads to the lake will not change between pre-implementation and post-implementation conditions. The goal for the monitoring design would be to test the null hypothesis and, if the null hypothesis is rejected, to conclude with some level of confidence that a change occurred. Monitoring the variables and the locations where a response is anticipated and monitoring close to the impaired or treated area will often help collect the data necessary to test a hypothesis.

If change is documented, it is usually desirable to be able to attribute the response to the implementation of management measures. This underscores the need for land treatment monitoring that provides data on the "cause" to relate to the water quality "effect." A true cause-effect relationship is difficult to document in a watershed project, but the results from well-designed monitoring in a watershed where implementation is properly targeted and fully carried out often support the argument that the management measures caused the change. The overall design of the monitoring program is critical to being able to make that argument.

Because the emphasis of this technical note is on assessing watershed project effectiveness, the types of designs described are as follows:

- Single-watershed before/after
- Above/below watersheds
- Side-by-side watersheds
- Paired watersheds
- Trend monitoring



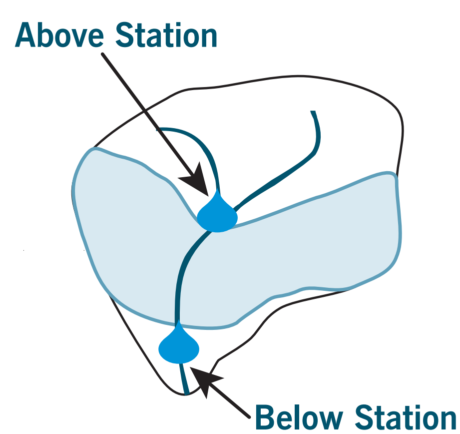
Single-Watershed Before/After

In this design a single monitoring station is typically placed at the downstream end of the project area and sampling is conducted before and after management measures are implemented. In most cases this design should be avoided (USDA, 1996).

The advantages of this design are that there is only one monitoring station, the design can be applied to most watersheds, and any data collected might provide a good database for trend analysis. The major disadvantage is that the effects of the watershed project are very difficult to separate from the effects of climate, hydrologic variation, and other factors. In short, it is hard, if not impossible, to attribute any water quality changes to a particular cause. This disadvantage makes this design unsuitable for watershed project evaluation. Paired t-tests are inappropriate for this design because there are no paired data points, but other statistical approaches, including a parametric t-test, might be feasible (USDA, 1996).

Above/Below Watersheds

In this design monitoring stations are placed above and below the area in which management measures are implemented. Also known as a nested design, this can be treated as a paired-watershed approach if monitoring is done before and after management measures are implemented (USDA, 1996). This design is not as

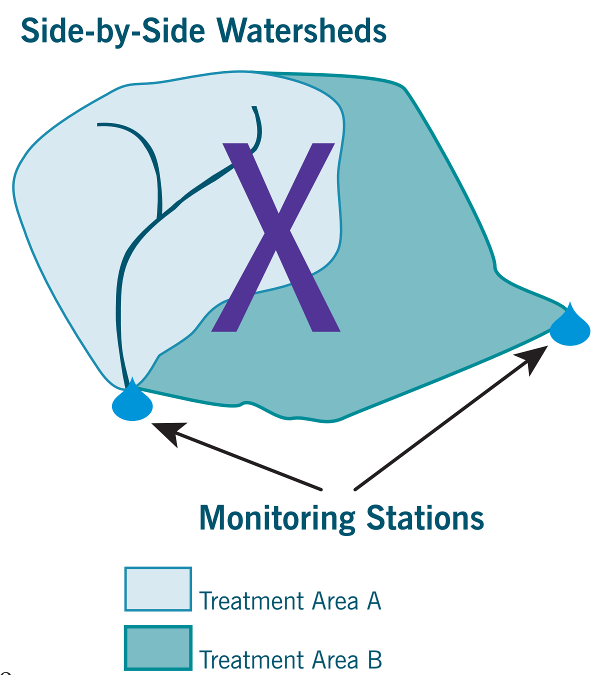


vulnerable to year-to-year climate variations as is single watershed monitoring; it is fairly easy to find a situation where treatment can be implemented between stations; and it is possible to attribute changes in water quality to specific causes, as long as management measures and other land use variables are monitored in both watersheds.

One disadvantage is that the water quality at the above station and that at the below station are not independent because upstream concentrations usually affect downstream concentrations. This might confound statistical analysis of the data. In addition, water quality differences between the monitoring stations might not be due solely to the management measures implemented. The characteristics of the watershed, such as upland versus downstream geology, could be quite different. For example, the headwaters area might be forested while the valley is heavily farmed, causing (or in response to) different soil conditions that might cause water quality differences. Note that these are problems associated with the above/below design if monitoring occurs *only after* implementation. If both pre-implementation and post-implementation monitoring are used, some of these problems can be addressed by a paired-watershed analysis. A paired t-test of the differences between paired above and below samples is appropriate for this design (USDA, 1996).

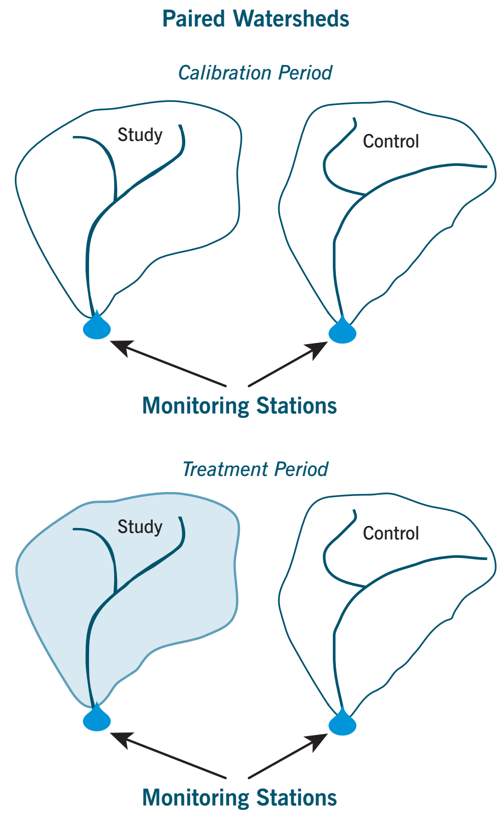
Side-by-Side Watersheds

In this design two adjacent or nearby watersheds are monitored independently. Land use or land management in the two watersheds is different. Pre-implementation monitoring is not conducted, and the two watersheds are not calibrated against each other (see Paired Watersheds). This design should be avoided because it offers no possibility to attribute cause to any measured changes in water quality. This design is essentially two single watersheds with only post-implementation monitoring in each.



A clear advantage of this design is that it is typically quite easy to find two watersheds with different land use or treatment. Although project personnel might feel good about monitoring two watersheds, they will find that the design is powerless to determine whether any measured differences in water quality between the two watersheds are the result of the practices rather than inherent differences in the characteristics of the two watersheds. Furthermore, the interactions between management measures and watersheds might differ as well because of differing watershed characteristics such as soil types, surface cover, and slopes. Statistical analysis for this design might not be appropriate, but it might be possible to use a paired t-test or nonparametric t-test of treatment means (USDA, 1996).

Paired Watersheds



In the paired-watershed design, there are two watersheds—control and study—and two monitoring periods—calibration and treatment (Clausen and Spooner, 1993). During the calibration period, the watersheds are treated identically (no changes preferably) and water quality data are collected from each watershed simultaneously. A statistical relationship (regression) is established between paired water quality data for the two watersheds. Treatment is then carried out in the study watershed during the treatment period, while the control watershed remains untreated. A new relationship between the paired data is then established after management practice implementation, and the change in this relationship is tested to indicate the effect of the treatment. Because practice implementation is not instantaneous, it is usually the case that there is a third period between calibration and treatment during which the practices are installed as needed. In some cases, water quality monitoring is discontinued during this construction period and picked up again after the practices are in place.

The advantages of this monitoring approach are that it controls for the effects of hydrologic variation and inherent watershed differences and it allows clear attribution of any measured water quality response to the management measures implemented, as long as the measures were implemented **only** in the study watershed. Some disadvantages are that it is twice as expensive as single watershed designs and takes two to three times longer to carry out. In addition, it is often difficult to find nearby watersheds suitable for this approach because of the need to control activities in both watersheds for the entire monitoring period. In agricultural settings it has often been the case that landowners in the control (no implementation) watershed are inspired to implement the same management measures as those in the study watershed, thereby violating the design. A post-monitoring implementation plan for the control watershed might solve this problem.

The basic statistical approach for this design is to develop significant linear regression relationships between the control watershed and treatment watershed for each water quality variable of interest during both the calibration and treatment periods (Clausen and Spooner, 1993; USDA, 1996). The significance of the relationship between paired observations is tested using analysis of variance, and the significance of the effect of the management measures is determined using analysis of covariance (Clausen and Spooner, 1993).

Trend Monitoring

Like the single-watershed before-and-after design, trend monitoring consists of a single monitoring station. The key difference, however, is that trend monitoring is intended to detect water quality changes that occur over a longer time frame (>10 years) due to the size and complexity of the watershed or more gradual implementation of management measures. It is important in trend monitoring to track land use and land management carefully to provide a basis for relating any measured water quality changes to treatment. Likewise, it is important to track precipitation and flow to help explain the variability in the monitoring variable(s) of interest. In an analysis of total phosphorus concentration, for example, a regression model could include both time and flow as model parameters. Examples of this type of regression analysis are abundant in the literature (see Spooner, Dickey, and Gilliam, 1990; Spooner, Jamieson, et al., 1987; USEPA, 2005b).

Trend Monitoring

Advantages of trend monitoring include the use of a single station and the relative ease with which a suitable watershed can be found. A key disadvantage is the need to commit to long-term water quality and land use/management monitoring because it is usually difficult to secure funding and capable staff over such a long period. It is also essential that data gaps be avoided, sampling and sample analysis methods remain the same throughout the monitoring period, and accurate hydrologic data be collected for the project duration (USDA, 1996). All long-term monitoring efforts run the risk of unwanted major land use changes occurring at some point, but the trend approach is particularly susceptible to this problem. Urban and developing areas are typically more likely than rural areas to experience such upheavals, but even agricultural areas can experience sudden changes in the types and numbers of livestock raised or the types of crops grown in response to market pressures or new regulations.

Scale

Watershed project monitoring can be carried out at the plot, field, and watershed scales (USDA, 1996). Plot-level monitoring is typically a short-duration (< 5 years) exercise designed to learn more about the fate and transport of pollutants or to quantify the effectiveness of specific management measures or practices. Well-designed plot-level monitoring is more akin to research than to watershed project effectiveness monitoring. Individual plots usually cover a small fraction of an acre. Although plot-level monitoring has notable advantages, such as replicates and controls, the results are not completely transferable to larger areas.

Monitoring at a scale of a few acres (e.g., an agricultural field) is often incorporated within an overall watershed project monitoring program to develop greater understanding of specific management measures or practices. Field-level monitoring is usually of short duration because treatment response can be measured more quickly than at the watershed scale. Such monitoring can be very effective in demonstrating and documenting the effectiveness of a management measure in a realistic setting. However, because edge-of-field results cannot be reliably extrapolated to a watershed scale, field-level monitoring is rarely sufficient for evaluating the effectiveness of a watershed project.

Watershed-level monitoring is the primary mechanism for evaluating the effectiveness of a watershed project because the projects are implemented at this scale. Watersheds can range in size from a few acres to several thousand acres depending on the project goals and setting. As a general rule, larger watersheds are slower to respond to treatment because of the pollutant transport mechanisms involved. This, however, might not be the case where biological monitoring is conducted, particularly in watersheds where habitat restoration is a key component of the implementation plan. Watersheds exceeding 5,000 acres or so generally require monitoring programs of 10 years or more to measure the effects of management measures, although the exact time frame depends on a range of factors, including the type of problem being addressed, the monitoring design employed, the weather during the monitoring period, and the type and extent of treatment implemented.

Variables

The basic monitoring program to assess watershed project effectiveness must focus on the pollutant sources identified in the watershed, the key pollutants from these sources, the water resources affected by these sources and pollutants, measures of designated use support in the affected water resources, and any biological and habitat issues of concern. The presence of existing monitoring efforts should not be a factor in determining the needs for the current watershed project, but coordination and data-sharing with ongoing efforts should be incorporated to the extent practicable to achieve the monitoring goals most efficiently. A number of tools are available to determine which water quality variables are most important for a particular project (USDA, 1996). It is important that variable selection be prioritized on the basis of the contribution of each variable to achieving the monitoring goals. Trade-offs among the number of variables, sample size, and number of monitoring sites should be considered based on both annual and long-term costs of sample collection and analysis because the relatively small annual cost savings associated with dropping some variables from the list might amount to considerable savings over the long term. These savings might be sufficient to increase sampling frequency, extend the project, or add monitoring sites to strengthen the monitoring program (i.e., to add options to the basic requirements).

Logistical factors, including site access and conditions, personnel availability, and travel times, also influence the selection of water quality variables to monitor. Some water quality constituents, such as soluble reactive phosphorus, have short (24-hour) holding times and require refrigeration while waiting for pickup. In contrast, samples for total phosphorus analysis can be held for as long as 7 days under proper conditions. Samples for bacteria analysis can rarely be collected by automated sampling equipment because of sterility issues and have even shorter allowable holding times (6 hours) between collection and analysis. Trade-offs between the expense and personnel effort to accommodate such constraints and the value of the resulting information must be considered when selecting variables to monitor.

In the end, monitoring data will be used to make statements about the effectiveness of the watershed project. These statements, or inferences, can be made only if the target population (e.g., the stream discharge to a lake or the fish assemblage in a stream segment) is sampled using a suitable statistical sampling plan. Key indicators of effectiveness (e.g., pollutant concentration or load, measures of fish diversity or abundance) must be identified at the beginning of the monitoring effort and tracked appropriately during the monitoring period. If one imagines the target population (e.g, stream discharge to a lake) to be divided into population units (e.g., each discrete cubic foot of water passing a point on the stream), the sampling plan boils down to how many population units to sample, when to sample them, where and how to sample them, and how to analyze and report on the samples taken.

Precipitation, air temperature, and other weather variables are also typically tracked to aid in data analysis and interpretation. For example, the relationship between stream flow and precipitation amount or intensity might change due to implementation of management measures, particularly in urban settings. Absent precipitation data, analysts might be unable to demonstrate any project benefits if, for example, flows remain the same after management measure implementation because of increased precipitation during that period. Flow itself is often used to aid in interpreting pollutant concentration data. The equipment for and design of precipitation and flow monitoring stations are described in the literature (USDA, 1996).

Sample Type

The statistical design selected for the project determines the general location of sampling stations. Waterbody type is a key factor in determining the specific sampling location for any particular monitoring design (USDA, 1996; USEPA, 1997a). The specific depth and horizontal placement of a sample intake within a waterbody are determined by the population units of interest.

Samples should represent the population units from which they are derived. For example, if each cubic foot of stream discharge is sampled, it is important that the samples have the same, or about the same, characteristics as the portion of the cubic foot that is not collected for analysis. In completely mixed systems this is no problem, but in many cases it is important to determine the best depth and horizontal placement of sample intakes (or bottles) to obtain a representative sample. If load estimation is the goal, it is important that the variable concentration in the sample is the same as that in the unsampled volume. If the interest is oxygen concentration at a certain depth, that depth is where sampling should occur. For biological monitoring, it might be important to sample particular habitat types (e.g., riffles or pools) preferentially.

There are a variety of sample collection options, several of which are described in manuals and beyond the scope of this technical note (e.g., APHA, 1995; Davis, 2005; USDA, 1996; USGS, 1977). Sample preservation, acceptable holding times, and transport considerations should be incorporated into the quality assurance/quality control plan for the project.

Number of Samples

The frequency of sampling is greatly influenced by the system being studied (USEPA, 1997a). Relatively stable systems (e.g., lakes with long hydraulic residence times) require less frequent sampling to characterize general conditions, whereas rapidly changing systems (e.g., urban streams) require more frequent sampling. For many urban streams it might be desirable to sample every storm event, whereas weekly sampling might be sufficient for rural streams. Storm event samples are often composited into weekly samples for laboratory analysis. Monthly or quarterly sampling could be adequate for some lakes, whereas reservoirs with short hydraulic residence times might require weekly sampling.

The monitoring objectives also influence the appropriate sampling frequency (USDA, 1996; USEPA, 1997a). Monitoring of management measure effectiveness requires a greater sampling frequency than does trend monitoring. Trend monitoring, however, is typically carried out over a far longer period, bringing total sample counts near to those for effectiveness monitoring. Examples of sample size calculations are available in the literature (Clausen and Spooner, 1993; USDA, 1996; USEPA, 2002).

Two-Sample t-tests

For designs in which t-tests are performed on two means (e.g., above vs. below, study vs. control), the number of samples needed can be determined based on an estimate of the smallest meaningful difference between the two groups, a value that is set based on relevant experience. Specifically, when paired-watershed designs are employed, the following equation can be used to determine whether a sufficient number of samples have been taken to detect the smallest meaningful difference (Clausen and Spooner, 1993):

$$\frac{S^2_{yx}}{d^2} = \frac{n_1 n_2}{n_1 + n_2} \left\{ \frac{1}{F\left(1 + \frac{F}{n_1 + n_2 - 2}\right)} \right\}$$

where

S^2_{yx} is the estimated residual variance about the regression,

d^2 is the square of the smallest meaningful (worthwhile) difference,

n_1 and n_2 are the numbers of observations in the calibration and treatment periods, and

F is the table value (p = 0.05) for the variance ratio at 1 and n_1+n_2-3 degrees of freedom.

Trend Analysis

In monitoring programs designed to evaluate long-term water quality trends, the number of samples needed can be estimated by applying the concept of the minimum detectable change (MDC), which is the minimum change in the mean value of a water quality parameter over time that is considered to be statistically significant (Spooner, Jamieson, et al., 1987; USEPA, 2002, 2005c). In addition, to support trend analysis, a sufficient number of samples must be collected to adequately represent seasonality or other sources of variability evident in the data. Such patterns of variation in data can be evaluated by applying exploratory data analysis to existing monitoring data (USEPA, 2005a). In general, a monthly sampling frequency (or taking more frequent samples that can be aggregated to monthly median values, for example) is generally the minimum frequency for sampling streams or rivers in nonpoint source situations. Sampling programs to support trend analysis should operate continuously for the entire project period, using consistent methods, locations, and schedules.

Timing

The sampling time chosen depends on the monitoring goals, the target population, and the anticipated relationship between management measure implementation and measured water quality. If seasonal impacts are expected, sampling should occur during the identified season(s). To determine the general conditions of the water resource, sampling should occur throughout the year, perhaps by weekly sampling. Weekly sampling is a form of systematic sampling in which samples are evenly spaced over time after a random starting point is chosen (Gilbert, 1987). Systematic sampling is generally more manageable than simple random sampling or even random sampling within segments or strata, although any of these options can be used to gauge general conditions. However, time-based systematic sampling schedules should be reviewed carefully to minimize the possibility of bias by either over-sampling or entirely missing regularly occurring events such as point source discharges, cyclical land management activities, or seasonal patterns.

For nonpoint source load estimation, sampling must occur during high-flow events because that is typically when the greatest pollutant loads occur. This is a form of stratified random sampling because high-flow or storm events are typically defined in some manner that triggers instruments (or people with bottles) to take samples. The stratum is the range of flow events during which samples are taken. In most cases, however, all such events are sampled unless equipment fails or staff fail to make it to the site. Storm-event sampling is more challenging than systematic sampling because the personnel involved might not always be available.

Some variables such as dissolved oxygen and temperature display diurnal patterns, and sampling time should be based on the portion of the pattern that is of interest. Biological monitoring is often done quarterly or on some other seasonal basis. In some climates summer sampling is impossible because there is no flow, while winter sampling might be precluded by freezing of the water resource.

Load Estimation

For any given watershed, the best approach for estimating loads will be determined on the basis of the needs and characteristics of the watershed. The critical principle at work is that nonpoint source pollutant concentrations are usually highly correlated with flow (i.e., high concentrations occur during high flows). Load estimation calculations must account for this fundamental relationship. Some general rules of thumb should be considered (USDA, 1996; Richards, 1997):

- Accuracy and precision increase with increased frequency of sampling.
- Grab, point, or instantaneous samples might be insufficient to determine loads unless concentrations are correlated to discharge that is measured continuously.
- Depth-integrated and width-integrated grab samples can account for stratification in concentration with depth or horizontally across a stream, but suitability in load estimation still depends on correlation with discharge.
- Time-weighted composite samples are not sufficient for load estimation because they do not adequately reflect changes in discharge and concentration during the period over which samples are composited.
- Flow-weighted composite samples are well suited to load estimation but difficult to collect because a stage-discharge relationship is required and a “smart sampler” is needed to trigger sampling as a function of flow rate. Predicting sample size and number of bottles needed in variable flow conditions is sometimes difficult.
- Systematic sampling is at least as efficient as simple random sampling if the sampling interval is not equal to a multiple of any strong period of fluctuation in the sampled population (e.g., sampling weekly on the day when a particular pollutant is always at its peak level due to scheduling by a discharger).
- Stratified random sampling with most samples taken during periods of high flow can be of great importance in providing increased precision for a given number of samples.

Biological Monitoring

For biological or habitat assessment, the project objectives and size of the watershed or study area determine the location of the sampling sites (USEPA, 1996). If the objective is to establish a baseline before an activity (implementing a BMP or imposing an impact), multiple sites should be randomly selected throughout the watershed to include reaches above and below the planned impact. For assessing the impact of a nonpoint source, sampling sites should be located in the areas immediately above and below the origin of the impact. A variable distance from the impact (sometimes as much as 1 to 2 miles for large watersheds) is often appropriate.

If an above/below design is used (USEPA, 1997), upstream sites should be located far enough upstream to avoid any residual effects from downstream impacts. This is especially important when monitoring fish community status. Upstream-downstream interaction among fish communities might extend over a half-mile distance during the summer months. Reference sites are needed to establish benchmark conditions for biological monitoring (USEPA, 1996). State or tribal water quality agencies should be consulted regarding appropriate reference sites relevant to the characteristics of the study area. If methods comparable to those of the state or tribe are used, the established and calibrated reference condition can be used as a benchmark for expected ecological condition. Reference sites to be sampled should be selected from the least disturbed areas that reflect realistically attainable conditions in the study area. Data from sampling reference sites can be used to calculate an “achievable value” for the project area (USEPA, 1991).

The number of reference sites should be sufficient to characterize the natural variability and the community potential for the monitored waterbodies. Reference sites should be situated in the same ecoregion with similar habitat types, as well as land use, land surface form, natural vegetation potential, and soils similar to each other and to those of the project area. Most state water quality programs have already established bioregions using this information, which will help to partition the natural variability (USEPA, 1999). For biological and habitat monitoring, sampling frequency is not as important as sampling during the appropriate index period. Consult with the state agency for appropriate biological index periods. Because biological indicators integrate effects over time, episodic or variable nonpoint sources can be monitored with a well-planned biosurvey during an index period. Reference conditions have been developed during periods of the most stable biological potential, and assessments of impact can be made at this time.

Biological assessment programs normally include surveys of fish, benthic macro-invertebrates, periphyton, macrophytes, or a combination for lotic systems. The physical habitat structure is assessed in the context of the quality of the habitat to support an appropriate structure and function of the aquatic community as measured against a reference condition (USEPA, 1999). A habitat assessment is performed at the same time as a biological survey.

The frequency of toxicity testing will depend on the period of greatest impact and the duration of the impact. Testing for toxic constituents should include multiple water samples during the period of greatest runoff, discharge, or other input source.

Land Treatment

Land use and land management variables should be tracked closely as part of watershed project monitoring efforts, although such information is often looked at only as available or at a far coarser scale for trend monitoring. Land treatment tracking is necessary to determine whether the watershed plan is being implemented appropriately. At a minimum, tracking data should include the location of practices, the time at which they were installed and became operational, and whether they continue to be operated properly after implementation is complete. Structural practices like waste storage lagoons or sediment basins might be easy to see and count, but management activities are more difficult to monitor. How have nitrogen and phosphorus applications changed under nutrient management? Are riparian buffers filtering sheet flow, or is runoff channelized through the buffer area? Are contractors following their erosion and sediment control plans?

Sometimes such questions can be answered only by asking the landowners. Some agricultural watershed projects have had success in asking farmers to keep records of tillage, manure and fertilizer application, harvest, and other management activities. Several Vermont projects, for example, used log books and regular interviews conducted by local crop management consultants to gather such information (Meals 1990, 1992, 2001). In urban settings, public works staff can be valuable sources of information. For example, data were collected by a variety of means through several sources in Washington's National Nonpoint Source Monitoring Program project (Washington DOE, 2003). These sources included on-site sewage system reports from the Thurston County Environmental Health Division, land use classification information from the Thurston County Assessor's Office, and digitized land-based data from the Thurston GeoData Center. Aerial photography and windshield or foot surveys are also useful in tracking structural and management practices. It is important to monitor not only where implementation is occurring but also all areas and actions in the watershed that might contribute to nonpoint source loads.

Surprises can derail the best watershed plan. An accidental release from a waste storage facility, a truck spill, or the isolated actions of a single "bad actor" can have serious water quality consequences. If the source is not documented, such a release can also raise questions about the effectiveness of the management measures implemented in the watershed.

The result of a good land use/land treatment monitoring program is a database of independent variables that will help explain changes in water quality in the future. The ability to attribute water quality changes to the implementation program or to other factors will be critical in evaluating the effectiveness of the practices implemented and making midcourse corrections. Available guidance provides recommendations for tracking agricultural, forestry, and urban management measures; rather than simply taking a list of variables from published materials, however, it is important to give thoughtful consideration to the site-specific needs of each watershed project (USEPA, 1997b; 1997c; 2001b).

Handling Noise in the Data

A certain amount of noise in environmental data is inevitable because of natural variability, random sampling error, measurement bias, and gross errors and mistakes (Gilbert, 1987). Watersheds and natural systems show tremendous variability in time and space. For example, no two runoff events are the same because of the many factors that influence them, including season, pre-event moisture conditions, and land management activities. Random sampling errors are related to the fact that any two samples taken from the same population unit will likely differ somewhat; statistical analysis techniques address this source of variability. Measurement bias is a problem in which measurements are consistently high or low—something that can and should be addressed with a strong quality assurance/quality control plan. Gross errors and mistakes include such things as forgetting to ice samples, incorrectly identifying biota in the field, and using the wrong blank to calibrate an instrument. Again, a strong quality assurance/quality control plan can identify these problems, but not always in time to save the sample.

To cut through the noise to find the meaning or signal, a monitoring program to measure the effectiveness of a watershed project must be designed to include recognition of the variability described here. A number of obstacles and constraints must be overcome in most watershed monitoring programs, including the following:

- Lack of control over activities that affect water quality
- Hydrologic variation across seasons and between years
- Incremental change brought on by a land treatment program
- Lag time in the response of natural systems to change
- Surprises, disasters, and other unusual events

Keeping these issues in mind during the design and operation of a monitoring program will increase the probability of success.

Quality Assurance and Quality Control Plan

Quality assurance and quality control practices should be an integral part of the development, design, and implementation of a watershed evaluation project to minimize or eliminate problems associated with the methodologies, data quality, and coordination of sampling and analysis efforts (USEPA, 1997a). Quality assurance activities are managerial in nature and include assigning roles and responsibilities to project staff, training staff, developing data quality objectives, validating data, and performing laboratory audits. Quality control procedures include the collection and analysis of blank, duplicate, and spiked samples and standard reference materials to ensure the integrity of analyses and regular inspection of equipment to ensure proper operation. USEPA implements its quality assurance/quality control program by requiring those who perform monitoring work for the Agency to prepare a Quality Management Plan and Quality Assurance Project Plan (QAPP) (USEPA, 1983; 1994). The process of preparing a QAPP and following its provisions goes a long way in ensuring correct and consistent data collection and a data set of known and acceptable quality (USEPA, 2001a).

Data Management, Analysis, and Reporting

While conducting a watershed evaluation project, it is important to document all data collected and used. All collected data should be validated with error checking, stored in a logically based and safe filing system with backups, and analyzed using proven approaches. Both hard and computerized copies of data should be maintained because each type of storage is susceptible to damage or loss. Both hard and computer copies should be housed separately from originals, and data should be backed up daily as long as new data are being acquired. A wide range of data analysis software packages are available for statistical analysis (USEPA, 1997a; 2005b).

A key element often overlooked or shortchanged in the design of a monitoring program is coordination and collaboration among the individuals and groups that are affected by or have a role in the monitoring program or watershed project. These stakeholders should be encouraged to participate in the design of the monitoring program regardless of their knowledge of monitoring because their support will be needed. In addition, periodic reporting to stakeholders is recommended to keep them involved in the project and to benefit from their feedback. Quarterly reports, although potentially time-consuming, are beneficial because they provide a strong incentive to stay on top of the data. Analysts can miss important project developments (e.g., radical changes in water quality due to sudden, unplanned land management changes or systematic lab errors) if they allow long delays before they analyze newly collected data; quarterly reporting reduces the risk that such events are overlooked.

References and Recommended Texts

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

Clausen, J.C., and J. Spooner. 1993. *Paired watershed study design*. EPA 841-F-93-009. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Davis, B.E. 2005. A guide to the proper selection and use of federally approved sediment and water-quality samplers. Open-file report 2005-1087. U.S. Geological Survey, Vicksburg, MS.

Gilbert, R.O. 1987. *Statistical methods for environmental pollution monitoring.* Van Nostrand Reinhold Company, New York.

ITFM (Intergovernmental Task Force on Monitoring Water Quality). 1995. *The strategy for improving water-quality monitoring in the United States.* Technical appendixes. Final report of the Intergovernmental Task Force on Monitoring Water Quality. Intergovernmental Task Force on Monitoring Water Quality, Washington, DC. February.

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.

Meals, D.W. 1990. *La Platte River Watershed Water Quality Monitoring and Analysis Program: Comprehensive final report*. Program report no. 12. Vermont Water Resources Research Center, University of Vermont, Burlington.

Meals, D.W. 1992. Relating land use and water quality in the St. Albans Bay watershed, Vermont. In *The National Rural Clean Water Program Symposium, September, 1992, Orlando, FL*, pp. 131–144. EPA/625/R-92/006. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

Meals, D.W. 2001. *Lake Champlain Basin Agricultural Watersheds Section 319 National Monitoring Program Project, Final Project Report: May, 1994–September, 2000.* Vermont Department of Environmental Conservation, Waterbury, VT.

Richards, R.P. 1997. *Estimation of pollutant loads in rivers and streams: A guidance document for NPS programs*. Draft. Water Quality Laboratory, Heidelberg College, Tiffin, OH.

Spooner, J., D.A. Dickey, and J.W. Gilliam. 1990. Determining and increasing the statistical sensitivity of nonpoint source control grab sample monitoring programs. In *Proceedings: Design of Water Quality Information Systems*, pp. 119–135. Information series no. 61. Colorado Water Resources Research Institute, Fort Collins, CO.

Spooner, J., C.J. Jamieson, R.P. Maas, and M.D. Smolen. 1987. Determining statistically significant changes in water pollutant concentrations. *Lake and Reservoir Management* 3: 195–201.

USDA (U.S. Department of Agriculture). 1996. *National handbook of water quality monitoring, part 600 national water quality handbook*. U.S. Department of Agriculture, Natural Resources Conservation Service, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1983. *Interim guidelines and specifications for preparing quality assurance project plans*. EPA-600/4-83-004. QAMS-005/80. U.S. Environmental Protection Agency, Office of Research and Development, Office of Monitoring Systems and Quality Assurance, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1991. *Watershed monitoring and reporting for section 319 national monitoring program projects*. U.S. Environmental Protection Agency, Office of Water, Office of Wetlands, Oceans, and Watersheds, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1994. *EPA requirements for quality assurance project plans for environmental data operations*. Interim final. EPA QA/R-5. U.S. Environmental Protection Agency, Quality Assurance Management Staff, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1996. *Biological criteria: Technical guidance for streams and small rivers.* Authored by G. Gibson, M. Barbour, J. Stribling, J. Gerritsen, and J. Karr. EPA-822-B-96-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1997a. *Monitoring guidance for determining the effectiveness of nonpoint source controls*. EPA 841-B-96-004. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1997b. *Techniques for tracking, evaluating, and reporting the implementation of nonpoint source control measures: Agriculture*. EPA 841-B-97-010. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1997. *Biological criteria: Technical guidance for survey design and statistical evaluation of biosurvey data.* Authored by K. Reckhow and W. Warren-Hicks. EPA-822-B-97-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1999. *Rapid Bioassessment Protocols for Use in Wadeable Streams and Rivers: Periphyton, Benthic Macroinvertebrates, and Fish.* Authored by M. Barbour, J. Gerritsen, B. Snyder, and J. Stribling. EPA-841-B-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1997c. *Techniques for tracking, evaluating, and reporting the implementation of nonpoint source control measures: Forestry*. EPA 841-B-97-009. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2001a. *EPA requirements for quality assurance project plans*. EPA/240/B-01/003. U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2001b. *Techniques for tracking, evaluating, and reporting the implementation of nonpoint source control measures: Urban*. EPA 841-B-00-007. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2002. *Urban stormwater BMP performance monitoring: A guidance manual for meeting the national stormwater BMP database requirements*. EPA 821-B-02-001. GeoSyntec Consultants, Urban Drainage and Flood Control District, and Urban Water Resources Research Council of the American Society of Civil Engineers in cooperation with U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2005a. *Monitoring data: Exploring your data, the first step*. NNPSMP technical note 1. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2005b. *Statistical trend analysis*. NNPSMP technical note 3. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2005c. *Minimum detectable change*. NNPSMP technical note 4. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USGS (U.S. Geological Survey). 1977. *National handbook of recommended methods for water-data acquisition*. U.S. Geological Survey, Office of Water Data Coordination, Reston, VA.

Washington DOE (Washington State Department of Ecology). 2003. *Totten and Eld Inlets clean water projects: Final report*. Publication no. 03-03-010. Washington State Department of Ecology, Olympia, WA.