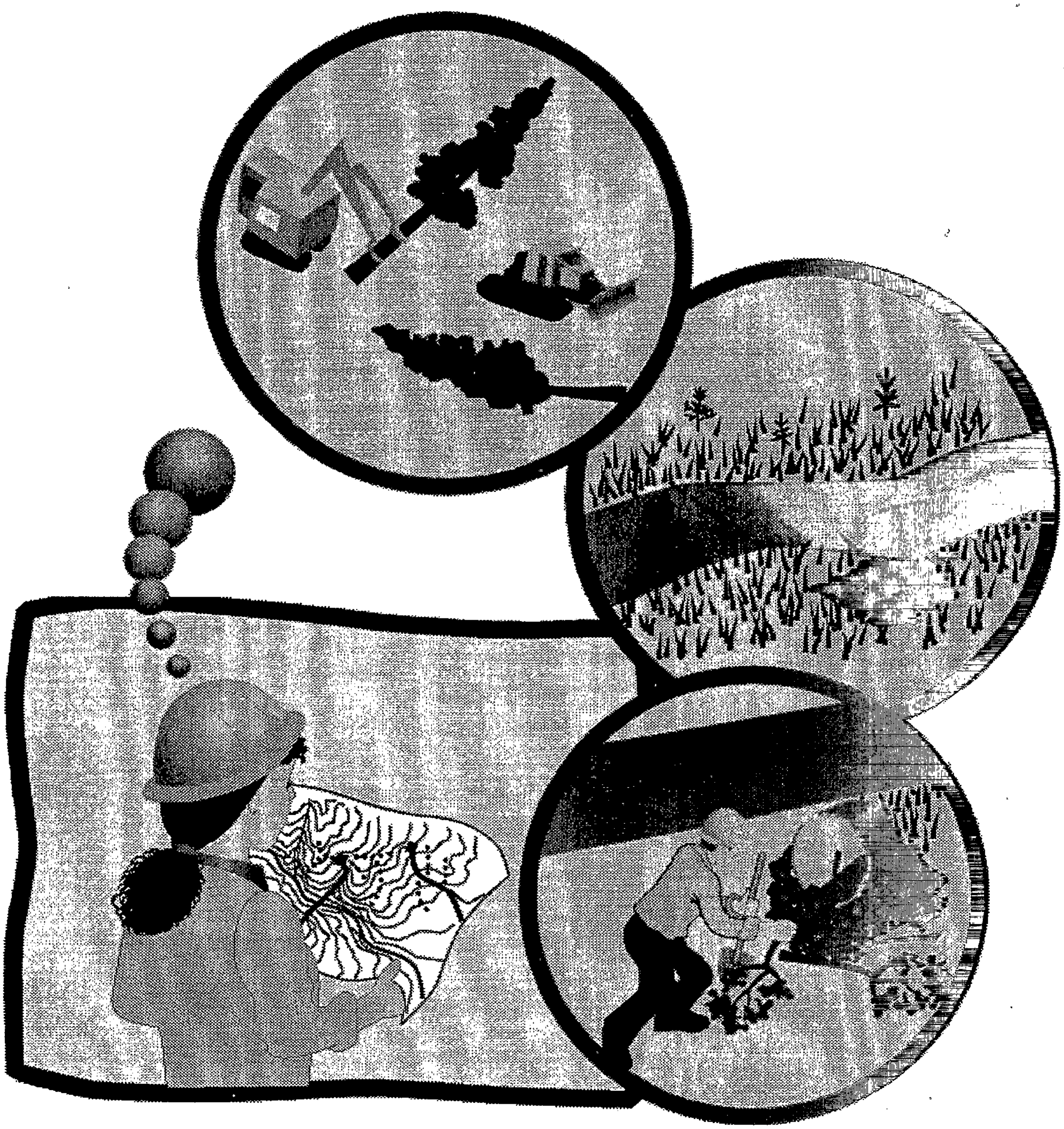
United States
Environmental
Protection Agency

Office of Water
(4503F)
Washington, DC

EPA 841-B-97-009
July 1997

# Techniques for Tracking, Evaluating, and Reporting The Implementation of Nonpoint Source Control Measures

## Forestry



# TECHNIQUES FOR TRACKING, EVALUATING, AND REPORTING THE IMPLEMENTATION OF NONPOINT SOURCE CONTROL MEASURES

## II. FORESTRY

**Final**
**July 1997**

Prepared for

Steve Dressing

Nonpoint Source Pollution Control Branch
United States Environmental Protection Agency

Prepared by

Tetra Tech, Inc.
EPA Contract No. 68-C3-0303
Work Assignment No. 4-51

# TABLE OF CONTENTS

# CHAPTER 1. INTRODUCTION

## 1.1 PURPOSE OF GUIDANCE

*The focus of this guide is on the design of monitoring programs to assess forestry management measure and best management practice implementation, with particular emphasis on statistical considerations.*

This guidance is intended to assist state, regional, and local environmental professionals in tracking the implementation of best management practices (BMPs) used to control nonpoint source pollution generated by forestry practices. Information is provided on methods for sample site selection, sample size estimation, sampling, and result evaluation and presentation. The focus of the guidance is on the statistical approaches needed to properly collect and analyze data that are accurate and defensible. A properly designed BMP implementation monitoring program can save both time and money. For example, in 1993 forestry operators notified the State of Idaho of 5,890 forestry operations (Colla, 1994). The cost of determining the status of BMP implementation on each of those forestry operations would have exceeded the amount budgeted, and thus statistical sampling of sites was needed. This document provides guidance for sampling representative forestry operations to yield summary statistics at a fraction of the cost of a comprehensive inventory.

Some forestry nonpoint source projects and programs combine BMP implementation monitoring with water quality monitoring to evaluate the effectiveness of BMPs in protecting water quality (Curtis et al., 1990; Rashin et al., 1994; USEPA, 1993b). For this type of monitoring to be successful, the scale of the project usually must be small (e.g., a watershed of a few hundred to a few thousand acres). Accurate records of all the sources of pollutants of concern and a census of how all BMPs are operating are very important for this type of monitoring effort. Otherwise, it can be extremely difficult to correlate BMP implementation with changes in stream water quality. This guidance does not address monitoring the implementation and effectiveness of all BMPs in a watershed. This guidance does provide information to help program managers gather statistically valid information to assess implementation of BMPs on a more general (e.g., statewide) basis. The benefits of implementation monitoring are presented in Section 1.3.

## 1.2 BACKGROUND

Pollution from nonpoint sources—sediment deposition, erosion, nutrients, contaminated runoff, hydrologic modifications that degrade water quality, and other diffuse sources of water pollution—is the largest cause of water quality impairment in the United States (USEPA, 1995). Congress passed the Coastal Zone Act Reauthorization Amendments of 1990 (CZARA) to help address nonpoint source pollution in coastal waters. CZARA provides that each state with an approved coastal zone management program develop and submit to the U.S. Environmental Protection Agency (EPA) and National Oceanic and Atmospheric Administration (NOAA) a Coastal Nonpoint Pollution Control Program (CNPCP). State programs must "provide for the implementation" of management measures in conformity with the EPA *Guidance Specifying Management Measures For Sources Of Nonpoint Pollution In Coastal Waters*, developed pursuant to Section 6217(g) of CZARA (USEPA, 1993a). Management measures (MMs), as defined in CZARA, are economically achievable measures to control the addition of pollutants to coastal waters, which reflect the greatest degree of pollutant reduction achievable through the application of the best available nonpoint pollution control practices, technologies, processes, siting criteria, operating methods, or other alternatives. Many of EPA's MMs are combinations of BMPs. For example, depending on site characteristics, implementation of the Road Management MM might involve use of the following BMPs: installing or regrading water bars; clearing road inlet and outlet ditches, catch basins, culverts, and road-crossing structures of obstructions; revegetating road surfaces; and inspecting closed roads.

CZARA does not specifically require that states monitor the implementation of MMs and BMPs as part of their CNPCPs. State CNPCPs must, however, provide for technical assistance to local governments and the public for implementing the MMs and BMPs. Section 6217(b) states:

> Each State program . . . shall provide for the implementation, at a minimum, of management measures . . . and shall also contain . . .
> (4) The provision of technical and other assistance to local governments and the public for implementing the measures . . . which may include assistance . . . to predict and assess the effectiveness of such measures . . . .

EPA and NOAA also have some responsibility under Section 6217 for providing technical assistance to implement state CNPCPs. Section 6217(d), Technical assistance, states:

> [NOAA and EPA] shall provide technical assistance . . . in developing and implementing programs. Such assistance shall include: . . .
> (4) methods to predict and assess the effects of coastal land use management measures on coastal water quality and designated uses.

This guidance document was developed to provide technical assistance as described in CZARA Sections 6217(b)(4) and 6217(d), but the techniques described can be used for other similar programs and projects. For instance, monitoring projects funded under Clean Water Act (CWA) Section 319(h) grants, efforts to implement total maximum daily loads developed under CWA Section 303(d), storm water permitting programs, and other programs could benefit from knowledge of BMP implementation.

Methods to assess the implementation of MMs and BMPs, then, are a key focus of the technical assistance to be provided by EPA and NOAA. Implementation assessments can be done on several scales. Site-specific assessments can be used to assess individual BMPs or MMs, and watershed assessments can be used to look at the cumulative effects of implementing multiple MMs. With regard to "site-specific" assessments, individual BMPs must be assessed at the appropriate scale for the BMP of interest. For example, to assess the implementation of MMs and BMPs for forest roads at harvest sites, only the roads at timber harvesting sites would need to be inspected. In this example, the scale would be a timber harvest area and the sites would be active and inactive roads at the harvest areas. To assess MM and BMP implementation at streamside management areas (SMAs), the proper scale might be a harvest area larger than 10 acres and the sites could be areas encompassed by buffer areas for 200-meter stretches of stream. For site preparation and forest regeneration, the scale and site might be an entire harvest site. Site-specific measurements can then be used to extrapolate to a watershed or statewide assessment. It is recognized that some studies might require a complete inventory of MM and BMP implementation across an entire watershed or other geographic area.

## 1.3 Types of Monitoring

The term *monitor* is defined as "to check or evaluate something on a constant or regular basis" (Academic Press, 1992). It is possible to distinguish among various types of monitoring. Two types, implementation and trend (i.e., trends in implementation) monitoring, are the focus of this guidance. These types of monitoring can be used to address the following goals:

- Determine the extent to which MMs and BMPs are being implemented in accordance with design standards and specifications.
- Determine whether there has been a change in the extent to which MMs and BMPs are being implemented.

In general, implementation monitoring is used to determine whether goals, objectives, standards, and management practices are being implemented as detailed in implementation plans. In the context of BMPs within state CNPCPs, implementation monitoring is used to determine the degree to which MMs and BMPs required or recommended by the CNPCPs are being implemented. If CNPCPs call for voluntary implementation of MMs and BMPs, implementation monitoring can be used to determine the success of the voluntary program (1) within a given monitoring period (e.g., 1 or 2 years); (2) during several monitoring periods, to determine any temporal trends in BMP implementation; or (3) in various regions of the state.

Trend monitoring involves long-term monitoring of changes in one or more parameters. As discussed in this guidance, public attitudes, land use, or the use of different forestry practices are examples of parameters that could be measured with trend monitoring. For example, the State of Idaho, Department of Lands, tracks trends in the number of forestry operations and enforcement actions (Colla, 1994). In addition, to isolate the impacts of MMs or BMPs on water quality, it is necessary to track their implementation over time.

Because trend monitoring involves measuring a change (or lack thereof) in some parameter over time, it is necessarily of longer duration than implementation monitoring and requires that a baseline, or starting point, be established. Any changes in the measured parameter are then detected in reference to the baseline.

Implementation and the related trend monitoring can be used to determine (1) which MMs and BMPs are being implemented, (2) whether MMs and BMPs are being implemented as designed, and (3) the need for increased efforts to promote or induce use of MMs and BMPs. Data from implementation monitoring, used in combination with other types of data (e.g., water quality data), can be useful in meeting a variety of other objectives, including the following (Hook et al., 1991; IDDHW, 1993; Schultz, 1992):

- To evaluate BMP effectiveness for protecting soil and water resources.
- To identify areas in need of further investigation.
- To establish a reference point of overall compliance with BMPs.
- To determine whether landowners/ forestry operators are aware of BMPs.
- To determine whether landowners/ forestry operators are using the advice of forestry BMP experts.
- To identify any BMP implementation problems specific to a land ownership category.
- To evaluate whether any forestry practices cause environmental damage.
- To compare the effectiveness of alternative BMPs.

MacDonald et al. (1991) describe additional types of monitoring, including effectiveness monitoring, baseline monitoring, project monitoring, validation monitoring, and compliance monitoring. As emphasized by MacDonald and others, these monitoring types are not mutually exclusive and the distinction between them is usually determined by the purpose of the monitoring.

Effectiveness monitoring is used to determine whether MMs or BMPs, as designed and implemented, are effective in meeting management goals and objectives. Effectiveness monitoring is a logical follow-up to implementation monitoring. It is essential that effectiveness monitoring include an assessment of the adequacy of the design and installation of MMs and BMPs. For instance, the objective of effectiveness monitoring could be to evaluate the effectiveness of MMs and BMPs *as designed and installed*, or to evaluate the effectiveness of MMs and BMPs *that are designed and installed adequately or to standards and specifications*. Effectiveness monitoring is the subject of another EPA guidance document, *Nonpoint Source Monitoring and Evaluation Guide* (USEPA, 1996).

Effectiveness monitoring for forestry BMPs is also addressed in a U.S. Forest Service document, *Evaluating the effectiveness of forestry best management practices in meeting water quality goals or standards* (Dissmeyer, 1994).

### 1.4 Quality Assurance and Quality Control

An integral part of the design phase of any nonpoint source pollution monitoring project is quality assurance and quality control (QA/QC). Development of a quality assurance project plan (QAPP) is the first step of incorporating QA/QC into a monitoring project. 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 QA/QC elements will be implemented throughout a project's life. It contains statements about the expectations and requirements of those for whom the data is being collected (i.e., the decision maker) 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 being well planned and executed. A thorough discussion of QA/QC is provided in Chapter 5 of EPA's *Nonpoint Source Monitoring and Evaluation Guide* (USEPA, 1996).

### 1.5 Data Management

Data management is a key component of a successful MM or BMP implementation monitoring effort. The data management system that is used—which includes the quality control and quality assurance aspects of data handling, how and where data are stored, and who manages the stored data—determines the reliability, longevity, and accessibility of the data. Provided that the data collection effort was planned and executed well, an organized and efficient data management system will ensure that the data can be used with confidence by those who must make decisions based upon it, the data will be useful as a baseline for similar data collection efforts in the future, the data will not become obsolete (or be misplaced!) quickly, and the data will be available to a variety of users for a variety of applications.

Serious consideration is often not given to a data management system prior to a data collection effort, which is precisely why it is so important to recognize the long-term value of a small investment of time and money in proper data management. Data management competes with other agency priorities for money, staff, and time, and if the importance and long-term value of proper data management is recognized early in a project's development, the more likely it will be to receive sufficient funding. Overall, data management might account for only a small portion of a project's total budget, but the return on the investment is great when it is considered that the larger investment in data collection can be rendered virtually useless unless data is managed adequately.

Two important aspects of data that should be considered when planning the initial data collection effort and a data management system are data life cycle and data accessibility. The data life cycle can be characterized by the following stages: (1) Data is collected; (2) data is checked for quality; (3) data is entered into a data base; (4) data is used, and (5) data eventually becomes obsolete. The expected usefulness and life span of the data should be considered during the initial stages of planning a data collection effort, when the money, staff, and time that are devoted to data collection must be weighed against its usefulness and longevity. Data with a limited use and that is likely to become obsolete soon after it is collected is a poorer investment decision than data with multiple applications and a long life span. If a data collection effort involves the collection of data of limited use and a short life span, it might be necessary to modify the data collection effort—either by changing its goals and objectives or by adding new ones—to increase the breadth and length of the data's applicability. A good data management system will ensure that any data that are collected will be useful for the greatest number of applications for the longest possible time.

Data accessibility is a critical factor in determining its usefulness. Data attains its highest value if it is as widely accessible as possible, if access to it requires the least amount of staff effort as possible, and if it can be used by others conveniently. If data are stored where those who might need it can obtain it with little assistance, it is more likely to be shared and used. The format for data storage determines how conveniently the data can be used. Electronic storage in a widely available and used data storage format makes it convenient to use. Storage as only a paper copy buried in a report, where any analysis requires entry into an electronic format or time-consuming manipulation, makes data extremely inconvenient to use and unlikely that it will be used.

The following should be considered for the development of a data management strategy:

- *What level of quality control should the data be subject to?* Data that will be used for a variety of purposes or that will be used for important decisions should receive a careful quality control check.

- *Where and how will the data be stored?* The options for data storage range from a printed final report on a bookshelf to an electronic data base accessible to government agencies and the public. Determining where and how data will be stored therefore also requires careful consideration of the question: *How accessible should the data be?*

- *Who will maintain the data base?* Data stored in a large data base might be managed by a professional data manager, while data kept in agency files might managed by people with various backgrounds over the course of time.

- *How much will data management cost?* As with all other aspects of a data collection effort, data management costs money and this cost must be balanced with all other costs involved in the project.

# CHAPTER 2. SAMPLING DESIGN

## 2.1. INTRODUCTION

This chapter discusses recommended methods for designing sampling programs to track and evaluate the implementation of nonpoint source control measures. This chapter does not address whether the management measures (MMs) or best management practices (BMPs) are effective since no water quality sampling is done. Because of the variation in forestry practices and related nonpoint source control measures implemented throughout the United States, the approaches taken by various states to track and evaluate nonpoint source control measure implementation will differ. Nevertheless, all approaches should be based on sound statistical methods for selecting sampling strategies, computing sample sizes, and evaluating data. EPA recommends that states consult with a trained statistician to be certain that the approach, design, and assumptions are appropriate to the task at hand.

As described in Chapter 1, implementation monitoring is the focus of this guidance. Effectiveness monitoring is the focus of another guidance prepared by EPA, the *Nonpoint Source Monitoring and Evaluation Guide* (USEPA, 1996). Dissmeyer (1994) also provides substantial information regarding QA/QC, statistical considerations, BMP effectiveness monitoring, and monitoring methods. The recommendations and examples in this chapter address two primary monitoring goals:

- Determine the extent to which MMs and BMPs are implemented in accordance with design standards and specifications.
- Determine whether there has been a change in the extent to which MMs and BMPs are being implemented.

For example, State forestry agencies might be interested in whether streamside management areas (SMAs) at harvest sites associated with all types of forest ownerships (industrial, private nonindustrial, federal, and state) are in compliance with design standards. State forestry agencies might also be interested in the percentage of owners of nonindustrial private forest land that are correctly implementing the BMPs specified in a voluntary implementation program.

### 2.1.1. Study Objectives

To develop a study design, clear, quantitative monitoring objectives must be developed. For example, the objective might be to estimate to within $\pm 5$ percent the percent of harvest sites that have adequate SMAs. Or perhaps a state is getting ready to implement new administrative procedures to ensure that purchasers of timber have been advised of applicable BMPs. In this case, detecting a 10 percent change in the number of operators that implement the BMPs specified in the timber sale contract might be of interest. In the first example, summary statistics are developed to describe the current status, whereas in the second example, some sort of statistical analysis (hypothesis testing) is performed to determine whether a significant change has really occurred. This choice has an impact on how the data are collected. As an example, summary statistics might require unbalanced sample allocations to account for variability such as site size, type, and ownership, whereas balanced designs (e.g., two sets of data with the same number of observations in each set) are more typical for hypothesis testing.

### 2.1.2. Probabilistic Sampling

Most study designs that are appropriate for tracking and evaluating implementation are based on a probabilistic approach since tracking every operator is not cost-effective. In a probabilistic approach, individuals are randomly selected from the entire group. The selected individuals are evaluated, and the results provide an unbiased assessment of the entire group. Applying the results from randomly selected individuals to the entire group is *statistical inference*. Statistical inference enables one to determine, for example, the probable percentage of timber sales with adequate SMAs without visiting every tract of land. One could also determine whether the change in timber sales with appropriate streamside management is within the range of what could occur by chance or the change is large enough to indicate a real modification of operator practices.

The group about which inferences are made is the population or *target population*, which consists of *population units*. The *sample population* is the set of population units that are directly available for measurement. For example, if the objective is to determine the degree to which adequate SMAs have been established, silvicultural operations for which SMAs are an appropriate BMP (e.g., timber sales with nearby streams) would be the sample population. Statistical inferences can be made only about the target population available for sampling. For example, if implementation of erosion control is being assessed and only public lands can be sampled, inferences cannot be made about the management of private lands.

The most common types of probabilistic sampling that can be used for implementation monitoring are summarized in Table 2-1. In general, probabilistic approaches are preferred. However, there might be circumstances under which targeted sampling should be used. Targeted sampling refers to using best professional judgement for selecting sample locations. For example, state foresters deciding to evaluate all timber sales in a given watershed would be targeted sampling. The choice of a sampling plan depends on study objectives, patterns of variability in the target population, cost-effectiveness of alternative plans, types of measurements to be made, and convenience (Gilbert, 1987).

*Simple random sampling* is the most basic type of sampling. Each unit of the target population has an equal chance of being selected. This type of sampling is appropriate when there are no major trends, cycles, or patterns in the target population (Cochran, 1977). Random sampling can be applied in a variety of ways including operator or timber sale selection. Random samples can also be taken at different times at a single site. Figure 2-1 provides an

Table 2-1. Applications of four sampling designs for implementation monitoring.

| Sampling Design | Comment |
|---|---|
| Simple Random Sampling | Each population unit has an equal probability of being selected. |
| Stratified Random Sampling | Useful when a sample population can be broken down into groups, or strata, that are internally more homogeneous than the entire sample population. Random samples are taken from each stratum although the probability of being selected might vary from stratum to stratum depending on cost and variability. |
| Cluster Sampling | Useful when there are a number of methods for defining population units and when individual units are clumped together. In this case, clusters are randomly selected and every unit in the cluster is measured. |
| Systematic Sampling | This sampling has a random starting point with each subsequent observation a fixed interval (space or time) from the previous observation. |

example of simple random sampling from a listing of harvest sites and from a map.

If the pattern of MM and BMP implementation is expected to be uniform across the state, simple random sampling is appropriate to estimate the extent of implementation. If, however, implementation is homogeneous only within certain categories (e.g., federal, state, or private lands), stratified random sampling should be used.

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. Simple random sampling is then used within each stratum. Stratification involves the use of categorical variables to group observations into more units, thereby reducing the variability of observations within each unit. For example, in a state with federal, state, and private forests, there might be different patterns of BMP implementation. Lands in the state could be divided into federal, state, and private as separate strata from which samples would be taken. In general, a larger number of samples should be taken in a stratum if the stratum is more variable, larger, or less costly to sample than other strata. For example, if BMP implementation is more variable on private lands, a greater number of sampling sites might be needed in that stratum to increase the precision of the overall estimate. Cochran (1977) found that stratified random sampling provides a better estimate of the mean for a population with a trend, followed in order by systematic sampling (discussed later) and simple random sampling. He also noted that stratification typically results in a smaller variance for the estimated mean or total than that which results from comparable simple random sampling.

| Harvest Site No. | Water Type | Ownership | County Code |
|---|---|---|---|
| 1 | Stream | Industry | 14 |
| 2 | Stream | Private Non-industrial | 3 |
| 3 | Lake/Pond | Industry | 12 |
| 4 | Stream | Industry | 11 |
| 5 | Lake/Pond | Private Non-industrial | 7 |
| 6 | Lake/Pond | Private Non-industrial | 4 |
| 7 | Lake/Pond | Industry | 7 |
| 8 | Stream | State | 10 |
| 9 | Stream | Federal | 10 |
| 10 | Lake/Pond | Private Non-industrial | 5 |
| 11 | Stream | Industry | 3 |
| 12 | Stream | State | 6 |
| ••• | ••• | ••• | ••• |
| 142 | Lake/Pond | Industry | 11 |
| 143 | Lake/Pond | State | 7 |
| 144 | Stream | State | 15 |
| 145 | Lake/Pond | Industry | 15 |
| 146 | Stream | Private Non-industrial | 8 |
| 147 | Stream | Private Non-industrial | 9 |
| 148 | Lake/Pond | Industry | 1 |
| 149 | Stream | Industry | 12 |
| 150 | Lake/Pond | Federal | 11 |

Figure 2-1a. Simple random sampling from a listing of harvest sites. In this listing, all harvest sites are presented as a single list and sites are selected randomly from the entire list. Shaded harvest sites represent those selected for sampling.

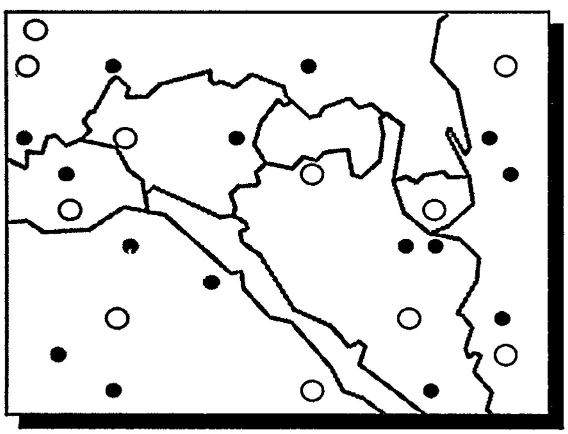


Figure 2-1b. Simple random sampling from a map. Dots represent harvest sites. All harvest sites of interest are represented on the map, and the sites to be sampled (open dots—○) were selected randomly from all harvest sites on the map. The shaded lines on the map could represent county, watershed, hydrologic, or some other boundary, but they are ignored for the purposes of simple random sampling.

If the state believes that there will be a difference between two or more subsets of the sites, such as between types of ownership or region, the sites can first be stratified into these subsets and a random sample taken within each subset (McNew, 1990). States with silviculture implementation monitoring programs commonly divide the sites by ownership and county and/or region before selecting survey sites. The goal of stratification is to increase the accuracy of the estimated mean values over what could have been obtained using simple random sampling of the entire population. The method makes use of prior information to divide the target population into subgroups that are internally homogeneous. There are a number of ways to "select" sites, or sets of sites, to be certain that important information will not be lost, or that MM or BMP use will not be misrepresented as a result of treating all potential survey sites as equal. Figure 2-2 provides an example of stratified random sampling from a listing of harvest sites and from a map.

Where data are available, it might be useful to compare the relative percentages of harvested timberland that is classified as having high, medium, and low erosion potentials. In cases where sediment is impacting water quality, highly erodible land might be responsible for a larger share of sediment delivery and would therefore be an important target for tracking the implementation of erosion controls. A stratified random sampling procedure could be used to estimate the percentage of total harvested timberland with different erosion potentials that have erosion controls in place. For other water quality problems (e.g., spawning habitat in decline), other stratification parameters (e.g., stream classification) might be more appropriate.

*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 the subdivisions are randomly selected for sampling. In one-stage cluster sampling, the selected clusters are sampled totally. In two-stage cluster sampling, random sampling is performed within each cluster (Gaugush, 1987). For example, this approach might be useful if a state wants to estimate the proportion of harvest sites that are following state-approved MMs or BMPs. All harvest sites in a particular county can be regarded as a single cluster. Once all clusters have been identified, specific clusters can be randomly chosen for sampling. 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 (1987) believes that the difficulty associated with analyzing cluster samples is compensated for by the reduced sampling requirements and cost. Figure 2-3 provides an example of cluster sampling from a listing of harvest sites and from a map.

*Systematic sampling* is used extensively in water quality monitoring programs because it is relatively easy to do from a management perspective. In systematic sampling the first sample has a random starting point and each

| Harvest Site No. | Water Type | Ownership | County Code |
|---|---|---|---|
| 1 | Stream | Industry | 14 |
| 3 | Lake/Pond | Industry | 12 |
| 4 | Stream | Industry | 11 |
| • • • | • • • | • • • | • • • |
| 148 | Lake/Pond | Industry | 1 |
| 149 | Stream | Industry | 12 |
| 2 | Stream | Private Non-industrial | 3 |
| 5 | Lake/Pond | Private Non-industrial | 7 |
| 6 | Lake/Pond | Private Non-industrial | 4 |
| • • • | • • • | • • • | • • • |
| 146 | Stream | Private Non-industrial | 8 |
| 147 | Stream | Private Non-industrial | 9 |
| 9 | Stream | Federal | 10 |
| 21 | Stream | Federal | 3 |
| • • • | • • • | • • • | • • • |
| 150 | Lake/Pond | Federal | 11 |
| 8 | Stream | State | 10 |
| 12 | Stream | State | 6 |
| • • • | • • • | • • • | • • • |
| 143 | Lake/Pond | State | 7 |
| 144 | Stream | State | 15 |

Figure 2-2a. Stratified random sampling from a listing of harvest sites. Within the listing, harvest sites were subdivided by the type of ownership. Then, considering only the private ownership category, harvest sites were selected randomly from within the category. The process of random site selection was then repeated separately for each ownership category. Shaded harvest sites within each category represent those selected for sampling.

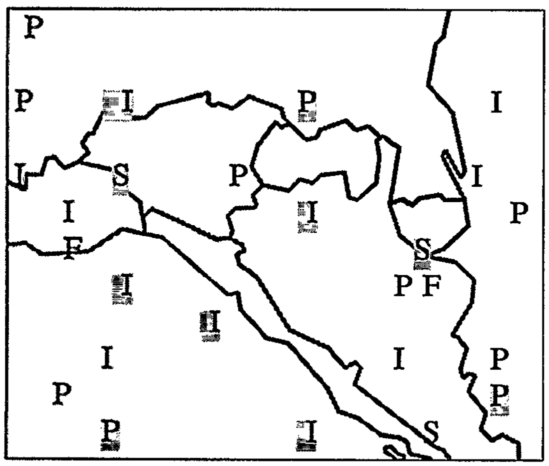


Figure 2-2b. Stratified random sampling from a map. Letters represent harvest sites, subdivided by type of ownership (P = private nonindustrial, I = industrial, F = federal, S = state). All harvest sites of interest are represented on the map. From all of the sites in one ownership category, sites were 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 sites into categories for the sampling process.

| Harvest Site No. | Water Type | Ownership | County Code |
|---|---|---|---|
| 148 | Lake/Pond | Industry | 1 |
| 2 | Stream | Private Non-industrial | 3 |
| 11 | Stream | Industry | 3 |
| 6 | Lake/Pond | Private Non-industrial | 4 |
| 10 | Lake/Pond | Private Non-industrial | 5 |
| 12 | Stream | State | 6 |
| 5 | Lake/Pond | Private Non-industrial | 7 |
| 7 | Lake/Pond | Industry | 7 |
| 143 | Lake/Pond | State | 7 |
| 146 | Stream | Private Non-industrial | 8 |
| 147 | Stream | Private Non-industrial | 9 |
| 8 | Stream | State | 10 |
| 9 | Stream | Federal | 10 |
| 4 | Stream | Industry | 11 |
| 142 | Lake/Pond | Industry | 11 |
| 150 | Lake/Pond | Federal | 11 |
| 3 | Lake/Pond | Industry | 12 |
| 149 | Stream | Industry | 12 |
| 1 | Stream | Industry | 14 |
| 144 | Stream | State | 15 |
| 145 | Lake/Pond | Industry | 15 |

Figure 2-3a. One-stage cluster sampling from a listing of harvest sites. Within the listing, harvest sties were subdivided by the county in which they were located. Some of these counties were then randomly selected, and all harvest sites within the selected counties were chosen for sampling. Shaded harvest sites represent those located in the counties selected (i.e., counties 3, 5, 8, 9, 11, and 15).

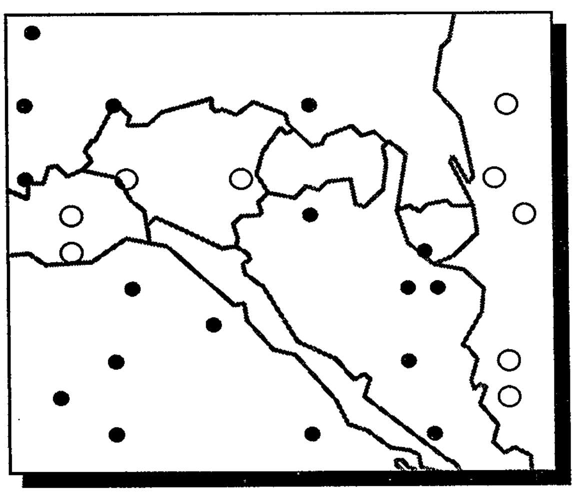


Figure 2-3b. Cluster sampling from a map. All harvest sites in the area of interest are represented on the map (closed {●} and open {○} dots). The shaded lines on the map represent county boundaries. Some of the counties were randomly selected, and all harvest sites within those counties (open dots - ○) 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.

subsequent sample has a constant distance from the previous sample. For example, if a sample size of 70 is desired from a mailing list of 700 operators, the first sample would be randomly selected from among the first 10 people, say the seventh person. Subsequent samples would then be based on the 17$^{th}$, 27$^{th}$, ..., 697$^{th}$ person. In comparison, a stratified random sampling approach might be to sort the mailing list by county and then to randomly select operators from each county. Figure 2-4 provides an example of systematic sampling from a listing of harvest sites and from a map.

In general, systematic sampling is superior to stratified random sampling when only one or two samples per stratum are taken for estimating the mean (Cochran, 1977) or when there is a known pattern of management measure implementation. 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. Cochran (1977) notes that on the average, simple random sampling and systematic sampling have equal variances. However, Cochran (1977) also states that for any single population for which the number of sampling units is small, the variance from systematic sampling is erratic and might 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 might 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 whether the randomly selected starting point is likely to have an impact on the results (Cochran, 1977).

Gilbert (1987) notes that assumptions about the population are required in estimating population variance from a single systematic sample of a given size. There are, however, systematic sampling approaches that do support unbiased estimation of population variance. They include 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.

### 2.1.3. Measurement and Sampling Errors

In addition to making sure that samples are representative of the sample population, it is also necessary to consider the types of bias or error that might be introduced into the study. *Measurement error* is the deviation of a measurement from the true value (e.g., the percent compliance with SMA specifications was estimated as 23 percent and the true value was 26 percent). A consistent under- or overestimation of the true value is referred to as *measurement bias*. Random

| Harvest Site No. | Water Type | Ownership | County Code |
|---|---|---|---|
| 1 | Stream | Industry | 14 |
| 2 | Stream | Private Non-industrial | 3 |
| 3 | Lake/Pond | Industry | 12 |
| 4 | Stream | Industry | 11 |
| 5 | Lake/Pond | Private Non-industrial | 7 |
| 6 | Lake/Pond | Private Non-industrial | 4 |
| 7 | Lake/Pond | Industry | 7 |
| 8 | Stream | State | 10 |
| 9 | Stream | Federal | 10 |
| 10 | Lake/Pond | Private Non-industrial | 5 |
| 11 | Stream | Industry | 3 |
| 12 | Stream | State | 6 |
| ••• | ••• | ••• | ••• |
| 142 | Lake/Pond | Industry | 11 |
| 143 | Lake/Pond | State | 7 |
| 144 | Stream | State | 15 |
| 145 | Lake/Pond | Industry | 15 |
| 146 | Stream | Private Non-industrial | 8 |
| 147 | Stream | Private Non-industrial | 9 |
| 148 | Lake/Pond | Industry | 1 |
| 149 | Stream | Industry | 12 |
| 150 | Lake/Pond | Federal | 11 |

Figure 2-4a. Systematic sampling from a listing of harvest sites. From a listing of all harvest sites of interest, an initial site (Harvest Site No. 2) was chosen randomly from among the first ten sites on the list. Every fifth site listed subsequently was then selected for sampling.

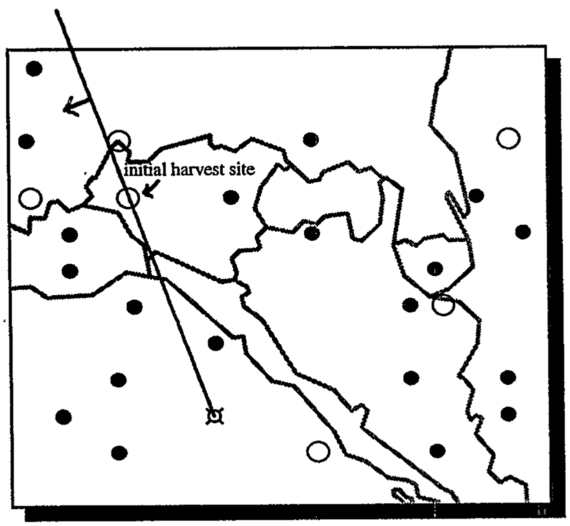



Figure 2-4b. Systematic sampling from a map. Dots (● and ○) represent harvest sites of interest. A single point on the map (¤) and one of the harvest sites were randomly selected. A line was stretched outward from the point to (and beyond) the selected harvest site. The line was then rotated about the 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 boundaries, soil type, watershed, or some other boundary, but were not used for the sampling. process.

*sampling error* arises from the variability from one population unit to the next (Gilbert, 1987), explaining why the proportion of operators using a certain BMP differs from one survey to another.

The goal of sampling is to obtain an accurate estimate by reducing the sampling and measurement errors to acceptable levels, while explaining as much of the variability as possible to improve the precision of the estimates (Gaugush, 1987). *Precision* is a measure of how close an agreement there is between individual measurements of the same population. The *accuracy* of a measurement refers to how close the measurement is to the true value. If a study has low bias and high precision, the results will have high accuracy. Figure 2-5 illustrates the relationship between bias, precision, and accuracy.

As suggested earlier, numerous sources of variability should be accounted for in developing a sampling design. Sampling errors are introduced by virtue of the natural variability within any given population of interest. Since sampling errors relate to MM or BMP implementation, the most effective method for reducing such errors is to carefully determine the target population and to stratify the target population to minimize the nonuniformity in each stratum.

Measurement errors can be minimized by ensuring that site inspections are well designed. If data are collected by sending staff out to inspect randomly selected harvest sites, the approach for inspecting the harvest sites should be consistent. For example, how do field personnel determine the percent of adequate SMAs, or what is the basis for determining whether a BMP has been properly implemented?

Reducing sampling errors below a certain point (relative to measurement errors) does not necessarily benefit the resulting analysis because total error is a function of the two types of errors. For example, if measurement errors such as response or interviewing errors are large, there is no point in taking a huge sample to reduce the sampling error of the estimate since the total error will be primarily determined by the measurement error. Measurement error is of particular concern when landowner surveys are used for implementation monitoring. Likewise, reducing measurement errors would not be worthwhile if only a small sample size were available for analysis because there would be a large sampling error (and therefore a large total error) regardless of the size of the measurement error. A proper balance between sampling and measurement errors should be maintained because research accuracy limits effective sample size and vice versa (Blalock, 1979).

### 2.1.4. Estimation and Hypothesis Testing

Rather than presenting every observation collected, the data analyst usually summarizes major characteristics with a few descriptive statistics. Descriptive statistics include any characteristic designed to summarize an important feature of a data set. A point estimate is a single number that represents the descriptive statistic. Statistics common to implementation monitoring

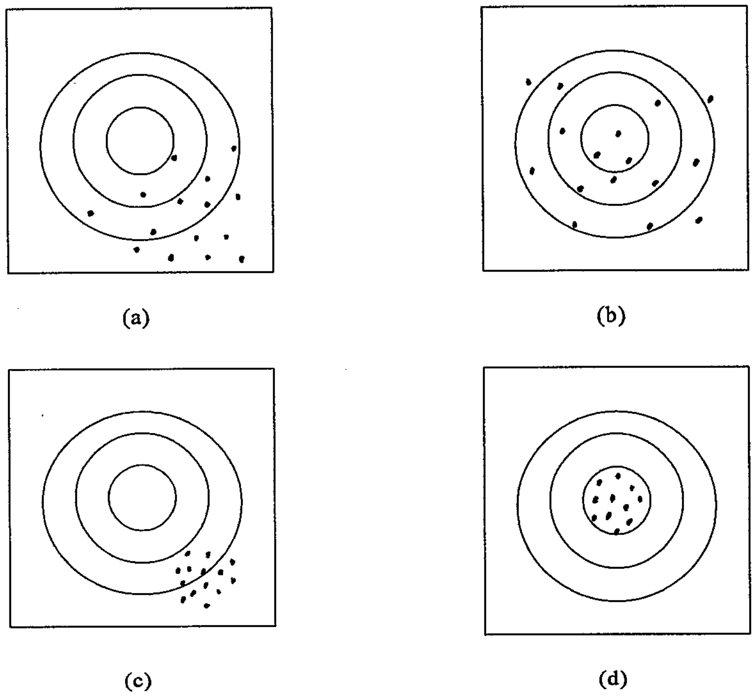



Figure 2-5. Graphical presentation of the relationship between bias, precision, and accuracy (after Gilbert, 1987). (a): high bias + low precision = low accuracy; (b): low bias + low precision = low accuracy; (c): high bias + high precision = low accuracy; and (d): low bias + high precision = high accuracy.

include proportions, means, medians, totals, and others. When estimating parameters of a population, such as the proportion or mean, it is useful to estimate the *confidence interval*. The confidence interval indicates the range in which the true value lies. For example, if it is estimated that 65 percent of waterbars on skid trails were installed in accordance with design standards and specifications and the 90 percent confidence limit is ±5 percent, there is a 90 percent chance that between 60 and 70 percent of the waterbars were installed correctly.

Hypothesis testing should be used to determine whether the level of MM and BMP implementation has changed over time. The *null hypothesis* *($H_o$)* is the root of hypothesis testing. Traditionally, $H_o$ is a statement of no change, no effect, or no difference; for example, "the proportion of properly installed waterbars after operator training is equal to the proportion of properly installed waterbars before operator training." The *alternative hypothesis* ($H_a$) is counter to $H_o$, traditionally being a statement of change, effect, or difference, for example. If $H_o$ is rejected, $H_a$ is accepted. Regardless of the statistical test selected for analyzing the data, the analyst must select the *significance level* ($\alpha$) of the test. That is, the analyst must determine what error level is acceptable based on the needs of decision makers. There are two types of errors in hypothesis testing:

Type I: $H_o$ is rejected when $H_o$ is really true.

Type II: $H_o$ is accepted when $H_o$ is really false.

Table 2-2 depicts these errors, with the magnitude of Type I errors represented by $\alpha$ and the magnitude of Type II errors represented by $\beta$. The probability of making a Type I error is equal to the $\alpha$ of the test and is selected by the data analyst. In most cases, managers or analysts will define *1-α* to be in the range of 0.90 to 0.99 (e.g., a confidence level of 90 to 99 percent), although there have been applications where *1-α* has been set to as low as 0.80. Selecting a 95 percent confidence level implies that the analyst will reject the $H_o$ when $H_o$ is true (i.e., a false positive) 5 percent of the time. The same notion applies to the confidence interval for point estimates described above: $\alpha$ is set to 0.10, and there is a 10 percent chance that the true percentage of properly installed waterbars is outside the 60 to 70 percent range. This implies that if the decisions to be made based on the analysis are major (i.e., affect many people in adverse or costly ways) the confidence level needs to be greater. For less significant decisions (i.e., low cost ramifications) the confidence level can be lower.

Type II error depends on the significance level, sample size, and variability, and which alternative hypothesis is true. *Power* ($1-\beta$) is defined as the probability of correctly rejecting $H_o$ when $H_o$ is false. In general, for a fixed sample size, $\alpha$ and $\beta$ vary inversely. For a fixed $\alpha$, $\beta$ can be reduced by increasing the sample size (Remington and Schork, 1970).

Table 2-2. Errors in hypothesis testing.

| | State of Affairs in the Population | |
|---|---|---|
| **Decision** | **$H_o$ is True** | **$H_o$ is False** |
| Accept $H_o$ | $1-\alpha$ (Confidence level) | $\beta$ (Type II error) |
| Reject $H_o$ | $\alpha$ (Significance level) (Type I error) | $1-\beta$ (Power) |

## 2.2. SAMPLING CONSIDERATIONS

In a document of this brevity, it is not possible to address all the issues that face technical staff who are responsible for developing and implementing studies to track and evaluate the implementation of nonpoint source control measures. For example, when is the best time to implement a survey or do on-site visits? In reality, it is difficult to pinpoint a single time of the year. Some BMPs can be checked any time of the year, whereas others have a small window of opportunity. If the goal of the study is to determine the effectiveness of an operator education program, sampling should be timed to ensure that there was sufficient time for outreach activities and for the operators to implement the desired practices. Furthermore, field personnel must have approval to perform a site visit on each tract of land to be sampled. Where access is denied, a randomly selected replacement site is needed.

### 2.2.1. Site Selection

From a study design perspective, all of these issues must be considered together when determining the sampling strategy. Site selection criteria will differ from state to state depending on the type of forestry practiced in the state, physical landscape, and intended purposes for the information obtained from the implementation monitoring. The following list indicates the typical site selection criteria culled from existing state implementation monitoring programs. (The corresponding state postal code is presented in parentheses.)

- *Site size:* minimum of 5 or 10 acres, depending on the region of the state (MN); minimum 10 acres (SC); minimum 5 acres (MT); minimum 20 acres (ID).

- *Proximity to a stream* (perennial or intermittent): within 300 feet of a stream, or a lake of at least 10 acres surface area (FL); within 200 feet of a stream (MT); sites did not have to be associated with streams or wetlands (SC); within 150 feet of a class II stream (ID).

- *Time of harvest:* within the past 1 year (SC); 1-3 years prior to the audit (MT); within 2 years of harvest (FL).

- *Site preparation:* only sites that had not been site prepared (SC); either slash piled and burned or waiting burning, or slash broadcast and scheduled to be burned (MT).

- *Volume harvested:* at least 7 MBF/ac (MT).

- *Compatibility with previous surveys:* sales had to meet the selection criteria of a previous study for comparability purposes (MT).

Other criteria that might be considered include erosion risk (e.g., more sampling sites could be placed in high-erosion-risk areas than in low-risk erosion areas) and beneficial use (bias sampling toward high use and/or sensitive areas).

### 2.2.2. Data to Support Site Selection

A list of harvest sites from which to choose those to be surveyed can be created from information obtained from timber harvesters. Depending on the state, the information is often in the form of harvest plans and timber sale contracts. These sources of information normally include:

- U.S. Forest Service offices.
- The state forestry agency, department, or division (for state lands and nonindustrial private).
- Private timber companies (Ehinger and Potts, 1991).

In addition, the Bureau of Land Management (BLM) manages a significant acreage of federal property and may have valuable information (IDDHW, 1993). Aerial photographs of the areas to be surveyed can be used to identify recent harvest sites as well, and this method of identifying the sites tends to remove site selection biases due to the distance of sites from roads or other forms of inconvenience that otherwise might make them less apt to be chosen for a survey.

The data necessary to select sites for BMP tracking will naturally depend on the site selection criteria. For instance, if sites must meet a minimum of board feet harvested, it will be necessary to know harvest volumes in order to select appropriate sites. The amount of data needed will increase as the number of site selection criteria increase, and this should be taken into account when deciding on the criteria, especially given the possibility that some of the data or types of data collected might be unavailable or unreliable.

### 2.2.3. Example State and Federal Programs

Several states and federal agencies have implemented programs for developing their implementation monitoring programs. This section describes those implemented by Florida, Montana, Idaho, and the USDA Forest Service.

#### 2.2.3.1. Florida

In Florida, the following site selection criteria are used (Vowell and Gilpin, 1994):

- All ownership classes are included.
- Only the northernmost 37 counties are included because most forestry activities occur in these counties.
- Timber harvesting, site preparation, tree planting, or some combination must have occurred within the past 2 years and within 300 feet of an intermittent or perennial stream or a lake 10 acres or larger.
- Each county has a predetermined number of survey sites based on the level of timber removal reported by the Forest Service.
- Sites are selected from fixed-wing aircraft using a random, predetermined flight pattern in each county.

County foresters randomly select qualifying sites along the flight pattern until they have located the number of survey sites assigned to their county.

This approach is a type of stratified random sampling. The entire population (entire state) is first divided into strata containing the northernmost 37 counties based on prior information that indicated most forestry operations occur in those counties. These strata are still too large to conduct random sampling; therefore, the criteria described above are used to reduce the strata to a manageable number given available resources.

#### 2.2.3.2. Montana and Idaho

Montana is interested in certain types of information related to BMP implementation, so they stratify their sample before selecting sites. They follow these steps:

- Information on the sites (e.g., ownership, erosion hazard) is compiled by watershed or basin.
- A list of all sales and information on them is compiled for each basin for the time period of interest (usually within 1-2 years of harvest date).
- Sites that do not meet the selection criteria are eliminated.
- Sites that do meet the selection criteria are ground-truthed.

This is a stratified random approach: Within drainage basins, sites are stratified first by ownership and then by erosion hazard (Ehinger and Potts, 1991; Schultz, 1992). Idaho also uses this approach, stratifying sites by geographic region and administrative category. This ensures that differences in MM and BMP implementation among different soil, geologic, and administrative groupings are not lost as would be the case if simple random sampling were used (IDDHW, 1993).

#### 2.2.3.3. U.S. Forest Service

The U.S. Forest Service (USDA, 1992) has developed a monitoring system for Region 5 of the Forest Service, Best Management Practice Evaluation Program (BMPEP), with the following objectives:

- Assess the degree of implementation of BMPs.
- Determine which BMPs are effective.
- Determine which BMPs need improvement or development.
- Fulfill Forest Land and Resource Management Plan BMP monitoring commitments.
- Provide a record of performance for management of nonpoint source pollution in Region 5 of the Forest Service.

These objectives are met through three evaluation phases: Administrative, on-site, and in-channel. In general, the first two phases deal with issues related to implementation monitoring, with administrative evaluation primarily addressing programmatic evaluation and on-site evaluation dealing primarily with individual practices. In-channel evaluation is an example of effectiveness monitoring.

In the BMPEP, forests are assigned the number and types of evaluations to be completed each year. To support statistical inference, the evaluations assigned to each forest must be performed at randomly identified sites. Sites to be evaluated are identified in two ways: Randomly and by selection ("selected" sites).

Randomly identified sites are essential for making statistical inferences regarding the implementation and effectiveness of BMPs. Random sites are picked from a pool of sites that meet specified criteria.

Selected sites are identified in various ways:

- Identified as part of a monitoring plan prescribed in an environmental assessment, environmental impact study, or land management plan.
- Identified as part of a Settlement of Negotiated Agreement.
- Part of a routine site visit.
- Follow-up evaluations upstream or In-channel Evaluation Sites, to discover sources of problems.
- Sites that are of particular interest to site administrators, specialists, and/or management due to their sensitivity, uniqueness, and other factors.
- Selected for a particular reason specific to local needs.

It is important to note that for statistical inference, the sample pool can only contain the randomly identified sites, not the "selected" sites. Selected sites must be clearly identified and kept separate from the random sites during data storage and analysis. Because on-site evaluation addresses a range of practices, corresponding methods are provided for developing sample pools for randomly selected sites. For example, the sample pool for SMAs should be developed using a Sale Area Map from the Pool of Timber Sales and counting the number of units that have designated SMAs. This constitutes the SMA sample pool.

The data obtained from the sources discussed above may not be precisely what are required by the state conducting the implementation monitoring survey. Ehinger and Potts (1991) report the following difficulties encountered in using data from the National Forest Service database:

- Flawed assumptions concerning the age of a harvest (some were found to be too old to meet the survey criteria).
- Uncertain age of roads.
- Units within 200 feet of stream on paper were in fact farther than 200 feet from a stream.

The following difficulties were associated with private nonindustrial forest units:

- Inadequate database to identify sales meeting the survey criteria.
- Permission to access landowner property not granted.

Landowners who did grant permission were interested primarily in demonstrating their BMP efforts to the state forest department, so this class of ownership was statistically biased.

On private industrial forest sites, a backlog of slash burnings on sale units was found, preventing their use (because of survey criteria), and state forests sites were mostly found to be farther than 200 feet from a stream, making them ineligible for the survey. States must be aware that these kinds of limitations will be encountered.

## 2.3. Sample Size Calculations

This section describes methods for estimating sample sizes to compute point estimates such as proportions and means, as well as detecting changes with a given significance level. Usually, several assumptions regarding data distribution, variability, and cost must be made to determine the sample size. Some assumptions might result in sample size estimates that are too high or too low. Depending on the sampling cost and cost for not sampling enough data, it must be decided whether to make conservative or "best-value" assumptions. Because the cost of visiting any individual site or group of sites is relatively constant, it is more economical to collect a few extra samples rather than realize later that additional data are needed. In most cases, the analyst should probably consider evaluating a range of assumptions regarding the impact of sample size and overall program cost. To maintain document brevity, some terms and definitions used in the remainder of this chapter are summarized in Table 2-3. These terms are consistent with those in most introductory-level statistics texts, and more information can be found there. Those with some statistical training will note that some of these definitions include an additional term referred to as the *finite population correction term* $(1-\phi)$, where $\phi$ is equal to $n/N$. In many applications, the number of population units in the sample population ($N$) is large in comparison to the number of population units sampled ($n$), and $(1-\phi)$ can be ignored. However, depending on the number of units (harvest sites for example) in a particular population, $N$ can become quite small. $N$ is determined by the definition of the sample population and the corresponding population units. If $\phi$ is greater than 0.1, the finite population correction factor should not be ignored (Cochran, 1977).

Applying any of the equations described in this section is difficult when no historical data set exists to quantify initial estimates of proportions, standard deviations, means, or coefficients of variation. To estimate these parameters, Cochran (1977) recommends four sources:

- Existing information on the same population or a similar population.

Table 2-3. Definitions used in sample size calculation equations.

| | | |
|---|---|---|
| N | = | total number of population units in sample population |
| n | = | number of samples |
| $n_0$ | = | preliminary estimate of sample size |
| a | = | number of successes |
| p | = | proportion of successes |
| q | = | proportion of failures (1-p) |
| $x_i$ | = | $i^{th}$ observation of a sample |
| $\bar{x}$ | = | sample mean |
| $s^2$ | = | sample variance |
| s | = | sample standard deviation |
| $N\bar{x}$ | = | total amount |
| μ | = | population mean |
| $\sigma^2$ | = | population variance |
| σ | = | population standard deviation |
| $C_v$ | = | coefficient of variation |
| $s^2(\bar{x})$ | = | variance of sample mean |
| φ | = | n/N (unless otherwise stated in text) |
| $s(\bar{x})$ | = | standard error (of sample mean) |
| 1-φ | = | finite population correction factor |
| d | = | allowable error |
| $d_r$ | = | relative error |
| $Z_\alpha$ | = | value corresponding to cumulative area of 1-α using the normal distribution (see Table A1). |
| $t_{\alpha,df}$ | = | value corresponding to cumulative area of 1-α using the student t distribution with df degrees of freedom (see Table A2). |

$$p = a/n \qquad q = 1-p$$

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

$$s = \sqrt{s^2} \qquad C_v = s/\bar{x}$$

$$d = |\bar{x}-\mu| \qquad d_r = \frac{|\bar{x}-\mu|}{\mu}$$

$$s^2(\bar{x}) = \frac{s^2}{n}(1-\phi) \qquad s(\bar{x}) = \frac{s}{\sqrt{n}}(1-\phi)^{0.5}$$

$$s(N\bar{x}) = \frac{Ns}{\sqrt{n}}(1-\phi)^{0.5} \qquad s(p) = \sqrt{\frac{pq}{n}(1-\phi)^{0.5}}$$

- 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 often the pilot sample is not representative of the entire population to be sampled.

- Informed judgment, or an educated guess.

It is important to note that this document only addresses estimating sample sizes with traditional parametric procedures. The methods described in this document should be appropriate in most cases, considering the type of data expected. If the data to be sampled are skewed, as with much water quality data, the analyst should plan to transform the data to something symmetric, if not normal, before computing sample sizes (Helsel and Hirsch, 1995). Kupper and Hafner (1989) also note that some of these equations tend to underestimate the necessary sample because power is not taken into consideration. Again, EPA recommends that if the analyst lacks a background in statistics, he/she should consult with a trained statistician to be certain that the approach, design, and assumptions are appropriate to the task at hand.

### 2.3.1. Simple Random Sampling

In simple random sampling, it is presumed that the sample population is relatively homogeneous and a difference in sampling costs or variability is not expected. If the cost or variability of any group within the sample population were different, it might be more appropriate to consider a stratified random sampling approach.

To estimate the proportion of harvest sites implementing a certain BMP or MM, such that the allowable error, $d$, meets the study precision requirements (i.e., the true proportion lies between $p$-$d$ and $p+d$ with a $1$-$\alpha$ confidence level), a preliminary estimate of sample size can be computed as (Snedecor and Cochran, 1980)

$$n_o = \frac{(Z_{1-\alpha/2})^2 p q}{d^2} \qquad \textbf{(2-1)}$$

*What sample size is necessary to estimate to within ±5 percent the proportion of harvest sites that have adequate SMAs?*

*What sample size is necessary to estimate the proportion of harvest sites that have adequate SMAs so that the relative error is less than 5 percent?*

If the proportion is expected to be a low number, using a constant allowable error might not be appropriate. Ten percent plus/minus 5 percent has a 50 percent relative error. Alternatively, the relative error, $d_r$, can be specified (i.e., the true proportion lies between $p-d_r p$ and $p+d_r p$ with a $1$-$\alpha$ confidence level) and a preliminary estimate of sample size can be computed as (Snedecor and Cochran, 1980)

$$n_o = \frac{(Z_{1-\alpha/2})^2 q}{d_r^2 p} \qquad \textbf{(2-2)}$$

In both equations, the analyst must make an initial estimate of $p$ before starting the study. In the first equation, a conservative sample size can be computed by assuming $p$ equal to 0.5. In the second equation the sample size gets larger as $p$ approaches 0 for constant $d_r$, and thus an informed initial estimate of $p$ is needed. Values of $\alpha$ typically range from 0.01 to 0.10. The final sample size is then estimated as (Snedecor and Cochran, 1980)

$$n = \begin{cases} \dfrac{n_0}{1+\phi} & \text{for } \phi > 0.1 \\ n_o & \text{otherwise} \end{cases} \qquad \textbf{(2-3)}$$

where $\phi$ is equal to $n_o/N$. Table 2-4 demonstrates the impact on $n$ of selecting $p$, $\alpha$, $d$, $d_r$, and $N$. For example, 278 random

Table 2-4. Comparison of sample size as a function of $p$, $\alpha$, $d$, $d_r$, and $N$ for estimating proportions using equations 2-1 through 2-3.

| Probability of Success, $p$ | Significance level, $\alpha$ | Allowable error, $d$ | Relative error, $d_r$ | Preliminary sample size, $n_o$ | Sample Size, $n$ — Number of Population Units in Sample Population, $N$: 500 | 750 | 1,000 | 2,000 | Large $N$ |
|---|---|---|---|---|---|---|---|---|---|
| 0.1 | 0.05 | 0.050 | 0.500 | 138 | 108 | 117 | 121 | 138 | 138 |
| 0.1 | 0.05 | 0.075 | 0.750 | 61 | 55 | 61 | 61 | 61 | 61 |
| 0.5 | 0.05 | 0.050 | 0.100 | 384 | 217 | 254 | 278 | 322 | 384 |
| 0.5 | 0.05 | 0.075 | 0.150 | 171 | 127 | 139 | 146 | 171 | 171 |
| 0.1 | 0.10 | 0.050 | 0.500 | 97 | 82 | 86 | 97 | 97 | 97 |
| 0.1 | 0.10 | 0.075 | 0.750 | 43 | 43 | 43 | 43 | 43 | 43 |
| 0.5 | 0.10 | 0.050 | 0.100 | 271 | 176 | 199 | 213 | 238 | 271 |
| 0.5 | 0.10 | 0.075 | 0.150 | 120 | 97 | 104 | 107 | 120 | 120 |

samples are needed to estimate the proportion of 1,000 harvest sites with adequate SMAs to within ±5 percent ($d$=0.05) with a 95 percent confidence level, assuming roughly one-half of harvest sites have adequate SMAs.

Suppose the goal is to estimate the average acreage per harvest site where erosion controls are used. The number of random samples required to achieve a desired margin of error when estimating the mean (i.e., the true mean lies between $\bar{x}-d$ and $\bar{x}+d$ with a $1-\alpha$ confidence level) is (Gilbert, 1987)

$$n = \frac{(t_{1-\alpha/2,n-1}s/d)^2}{1 + (t_{1-\alpha/2,n-1}s/d)^2/N} \qquad \textbf{(2-4)}$$

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

$$n = (t_{1-\alpha/2,n-1}s/d)^2 \qquad \textbf{(2-5)}$$

Since the Student's $t$ value is a function of $n$, Equations 2-4 and 2-5 are applied iteratively. That is, guess at what $n$ will be, look up $t_{1-\alpha/2,n-1}$ from Table A2, and compute a revised $n$. If the initial guess of $n$ and the revised $n$ are different, use the revised $n$ as the new guess, and repeat the process until the computed value of $n$ converges with the guessed value. If the population standard

*What sample size is necessary to estimate the average number of acres per harvest site using erosion controls to within ±25 acres?*

*What sample size is necessary to estimate the average number of acres per harvest site using erosion controls to within ±10 percent?*

deviation is known (not too likely), rather than estimated, the above equation can be further simplified to

$$n = (Z_{1-\alpha/2}\sigma/d)^2 \qquad \textbf{(2-6)}$$

To keep the relative error of the mean estimate below a certain level (i.e., the true mean lies between $\bar{x}-d_r\bar{x}$ and $\bar{x}+d_r\bar{x}$ with a $1-\alpha$ confidence level), the sample size can be computed with (Gilbert, 1987)

$$n = \frac{(t_{1-\alpha/2,n-1}C_v/d_r)^2}{1+(t_{1-\alpha/2,n-1}C_v/d_r)^2/N} \qquad \textbf{(2-7)}$$

$C_v$ is usually less variable from study to study than are estimates of the standard deviation, which are used in Equations 2-4 through 2-6. Professional judgment and experience, typically based on previous studies, are required to estimate $C_v$. Had $C_v$ been known, $Z_{1-\alpha/2}$ would have been used in place of $t_{1-\alpha/2,n-1}$ in Equation 2-7. If $N$ is large, Equation 2-7 simplifies to:

$$n = (t_{1-\alpha/2,n-1}C_v/d_r)^2 \qquad \textbf{(2-8)}$$

For Company X, harvest sites range in size from 20 to 400 acres although most are less than 80 acres in size. The goal of the sampling program is to estimate the average number of harvested acres using erosion controls. However, the investigator is concerned about skewing the mean estimate with the few large sites. As a result, the sample population for this analysis is the 430 harvested sites with less than 80 total acres. The investigator also wants to keep the relative error under 15 percent (i.e., $d_r$ = 0.15) with a 90 percent confidence level.

Unfortunately, this is the first study that Company X has done and there is no information about $C_v$ or $s$. The investigator, however, is familiar with a recent study done by another company. Based on that study, the investigator estimates the $C_v$ as 0.6 and $s$ equal to 30. As a first-cut approximation, Equation 2-6 was applied with $Z_{1-\alpha/2}$ equal to 1.645 and assuming $N$ is large:

$$n = (1.645 * 0.6/0.15)^2 = 43.3 \approx 44 \textit{ samples}$$

Since $n/N$ is greater than 0.1 and $C_v$ is estimated (i.e., not known), it is best to reestimate $n$ with Equation 2-7 using 44 samples as the initial guess of $n$. In this case, $t_{1-\alpha/2,n-1}$ is obtained from Table A2 as 1.6811.

$$n = \frac{(1.6811 \times 0.6/0.15)^2}{1+(1.6811 \times 0.6/0.15)^2/430} = 40.9 \approx 41 \textit{ samples}$$

Notice that the revised sample is somewhat smaller than the initial guess of $n$. In this case it is recommended to reapply the Equation 2-7 using 41 samples as the revised guess of $n$. In this case, $t_{1-\alpha/2,n-1}$ is obtained from Table A2 as 1.6839.

$$n = \frac{(1.6839 \times 0.6/0.15)^2}{1+(1.6839 \times 0.6/0.15)^2/430} = 41.0 \approx 41 \textit{ samples}$$

Since the revised sample size matches the estimated sample size on which $t_{1-\alpha/2,n-1}$ was based, no further iterations are necessary The proposed study should include 41 harvested sites randomly selected from the 430 sites with less than 80 total acres.

When interest is focused on whether the level of BMP implementation has changed, it is necessary to estimate the extent of implementation at two different time periods.

Alternatively, the proportion from two different populations can be compared. In either case, two independent random samples are taken and a hypothesis test is used to determine whether there has been a significant change in implementation. (See Snedecor and Cochran (1980) for sample size calculations for matched data.) Consider an example in which the proportion of waterbars that effectively divert water from the skid trail will be estimated at two time periods. What sample size is needed?

> *What sample size is necessary to determine whether there is a 20 percent difference in BMP implementation before and after an operator training program?*
>
> *What sample size is necessary to detect a 30-acre increase in average harvested acreage per site using erosion controls when comparing private and public timber sales?*

To compute sample sizes for comparing two proportions, $p_1$ and $p_2$, it is necessary to provide a best estimate for $p_1$ and $p_2$, as well as specifying the significance level and power ($1$-$\beta$). Recall that power is equal to the probability of rejecting $H_o$ when $H_o$ is false. Given this information, the analyst substitutes these values into (Snedecor and Cochran, 1980)

$$n_o = (Z_\alpha + Z_{2\beta})^2 \frac{(p_1 q_1 + p_2 q_2)}{(p_2 - p_1)^2} \qquad \textbf{(2-9)}$$

where $Z_\alpha$ and $Z_{2\beta}$ correspond to the normal deviate. Although this equation assumes that $N$ is large, it is acceptable for practical use (Snedecor and Cochran, 1980). Common values of $(Z_\alpha + Z_{2\beta})^2$ are summarized in Table 2-5. To account for $p_1$ and $p_2$ being estimated, $Z$ should be replaced with $t$. In lieu of an iterative calculation, Snedecor and Cochran (1980) propose the following approach: (1) compute $n_o$ using Equation 2-9; (2) round $n_o$ up to the next highest integer, $f$; and (3) multiply $n_o$ by $(f+3)/(f+1)$ to derive the final estimate of $n$.

To detect a difference in proportions of 0.20 with a two-sided test, $\alpha$ equal to 0.05, $1$-$\beta$ equal to 0.90, and an estimate of $p_1$ and $p_2$ equal to 0.4 and 0.6, $n_o$ is computed as

$$n_o = 10.51 \frac{[(0.4)(0.6) + (0.6)(0.4)]}{(0.6 - 0.4)^2} = 126.1$$

Rounding 126.1 to the next highest integer, $f$ is equal to 127, and $n$ is computed as 126.1 x 130/128 or 128.1. Therefore, 129 samples in each random sample, or 258 total samples, are needed to detect a difference in proportions of 0.2. Beware of other sources of information that give significantly lower estimates of sample size. In some cases the other sources do not specify $1$-$\beta$; in all cases, it is important that an "apples-to-apples" comparison is being made.

To compare the average from two random samples to detect a change of $\delta$ (i.e., $\bar{x}_2$-$\bar{x}_1$), the following equation is used:

$$n_o = (Z_\alpha + Z_{2\beta})^2 \frac{(s_1^2 + s_2^2)}{\delta^2} \qquad \textbf{(2-10)}$$

Common values of $(Z_\alpha + Z_{2\beta})^2$ are summarized in Table 2-5. To account for $s_1$ and $s_2$ being estimated, $Z$ should be replaced with $t$. In lieu of an iterative calculation, Snedecor and Cochran (1980) propose the

Table 2-5. Common values of $(Z_\alpha + Z_{2\beta})^2$ for estimating sample size for use with equations 2-9 and 2-10.

| Power, 1-β | α for One-sided Test | | | α for Two-sided Test | | |
|---|---|---|---|---|---|---|
| | 0.01 | 0.05 | 0.10 | 0.01 | 0.05 | 0.10 |
| 0.80 | 10.04 | 6.18 | 4.51 | 11.68 | 7.85 | 6.18 |
| 0.85 | 11.31 | 7.19 | 5.37 | 13.05 | 8.98 | 7.19 |
| 0.90 | 13.02 | 8.56 | 6.57 | 14.88 | 10.51 | 8.56 |
| 0.95 | 15.77 | 10.82 | 8.56 | 17.81 | 12.99 | 10.82 |
| 0.99 | 21.65 | 15.77 | 13.02 | 24.03 | 18.37 | 15.77 |

following approach: (1) compute $n_o$ using Equation 2-10; (2) round $n_o$ up to the next highest integer, $f$; and (3) multiply $n_o$ by $(f+3)/(f+1)$ to derive the final estimate of $n$.

Continuing the Company X example above, where $s$ was estimated as 30 acres, the investigator will also want to compare the average number of harvested acres that used erosion controls to the average number of harvested acres that used erosion controls in a few years. To demonstrate success, the investigator believes that it will be necessary to detect a 20-acre increase. Although the standard deviation might change after the operator training program, there is no particular reason to propose a different $s$ at this point. To detect a difference of 20 acres with a two-sided test, α equal to 0.05, $1$-β equal to 0.90, and an estimate of $s_1$ and $s_2$ equal to 30, $n_o$ is computed as

$$n_o = 10.51 \frac{(30^2 + 30^2)}{20^2} = 47.3 \qquad \textbf{(2-11)}$$

Rounding 47.3 to the next highest integer, $f$ is equal to 48, and $n$ is computed as (47.3)• (51/49) or 49.2. Therefore 50 samples in each random sample, or 100 total samples, are needed to detect a difference of 20 acres.

### *2.3.2. Stratified Random Sampling*

The key reason for selecting a stratified random sampling strategy over simple random sampling is to divide a heterogeneous population into more homogeneous groups. If populations are grouped based on size (e.g., site size) when there is a large number of small units and a few larger units, a large gain in precision can be expected (Snedecor and Cochran, 1980). Stratifying also allows the investigator to efficiently allocate sampling

*What sample size is necessary to estimate the average SMA width per harvest site when there is a wide variety of stream types and site conditions?*

resources based on cost. The stratum mean, $\bar{x}_h$, is computed using the standard approach for estimating the mean. The overall mean, $\bar{x}_{st}$, is computed as

$$\overline{x_{st}} = \sum_{h=1}^{L} W_h \overline{x_h} \tag{2-12}$$

where $L$ is the number of strata and $W_h$ is the relative size of the h$^{th}$ stratum. $W_h$ can be computed as $N_h/N$ where $N_h$ and $N$ are the number of population units in the h$^{th}$ stratum and the total number of population units across all strata, respectively. Assuming that simple random sampling is used within each stratum, the variance of $\bar{x}_{st}$ is estimated as (Gilbert, 1987)

$$s^2(\overline{x_{st}}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_h^2}{n_h} \tag{2-13}$$

where $n_h$ is the number of samples in the h$^{th}$ stratum and $s_h^2$ is computed as (Gilbert, 1987)

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_{h,i} - \overline{x_h})^2 \tag{2-14}$$

There are several procedures for computing sample sizes. The method described below allocates samples based on stratum size, variability, and unit sampling cost. If $s^2(\bar{x}_{st})$ is specified as $V$ for a design goal, $n$ can be obtained from (Gilbert, 1987)

$$n = \frac{\left(\sum_{h=1}^{L} W_h s_h \sqrt{c_h}\right) \sum_{h=1}^{L} W_h s_h / \sqrt{c_h}}{V + \frac{1}{N} \sum_{h=1}^{L} W_h s_h^2} \tag{2-15}$$

where $c_h$ is the per unit sampling cost in the h$^{th}$ stratum and $n_h$ is estimated as (Gilbert, 1987)

$$n_h = n \frac{W_h s_h / \sqrt{c_h}}{\sum_{h=1}^{L} W_h s_h / \sqrt{c_h}} \tag{2-16}$$

In the discussion above, the goal is to estimate an overall mean. To apply a stratified random sampling approach to estimating proportions, $p_h$, $p_{st}$, $p_h q_h$, and $s^2(p_{st})$ should be substituted for $\bar{x}_h$, $\bar{x}_{st}$, $s_h^2$, and $s^2(\bar{x}_{st})$ in the above equations, respectively.

To demonstrate the above approach, consider the Company X example again. In addition to the 430 sites that are less than 80 acres, there are 100 sites that range in size from 81 to 200 acres, 50 sites that range in size from 201 to 300 acres, and 20 sites that range in size from 301 to 400 acres. Table 2-6 presents three basic scenarios for estimating sample size. In the first scenario, $s_h$ and $c_h$ are assumed equal among all strata. Using a design goal of $V$ equal to 100 and applying Equation 2-15 yields a total sample size of 41.9 or 42. Since $s_h$ and $c_h$ are uniform, these samples are allocated proportionally to $W_h$, which is referred to as *proportional allocation*. This allocation can be verified by comparing the percent sample allocation to $W_h$. Due to rounding up, a total of 44 samples are allocated.

Under the second scenario, referred to as the *Neyman allocation*, the variability between strata changes, but unit sample cost is constant. In this example, $s_h$ increases by 15 between strata. Because of the increased

Table 2-6. Allocation of samples.

| Harvest Site Size (acres) | Number of Harvest Sites ($N_h$) | Relative Size ($W_h$) | Standard Deviation ($s_h$) | Unit Sample Cost ($c_h$) | Sample Allocation | |
|---|---|---|---|---|---|---|
| | | | | | Number | % |
| A) Proportional allocation ($s_h$ and $c_h$ are constant) | | | | | | |
| 20-80 | 430 | 0.7167 | 30 | 1 | 31 | 70.5 |
| 81-200 | 100 | 0.1667 | 30 | 1 | 7 | 15.9 |
| 201-300 | 50 | 0.0833 | 30 | 1 | 4 | 9.1 |
| 301-400 | 20 | 0.0333 | 30 | 1 | 2 | 4.5 |
| Using Equation 2-15, $n$ is equal to 41.9. Applying Equation 2-16 to each stratum yields a total of 44 samples after rounding up to the next integer. | | | | | | |
| B) Neyman allocation ($c_h$ is constant) | | | | | | |
| 20-80 | 430 | 0.7167 | 30 | 1 | 35 | 56.5 |
| 81-200 | 100 | 0.1667 | 45 | 1 | 13 | 21.0 |
| 201-300 | 50 | 0.0833 | 60 | 1 | 9 | 14.5 |
| 301-400 | 20 | 0.0333 | 75 | 1 | 5 | 8.1 |
| Using Equation 2-15, $n$ is equal to 59.3. Applying Equation 2-16 to each stratum yields a total of 62 samples after rounding up to the next integer. | | | | | | |
| C) Allocation where $s_h$ and $c_h$ are not constant | | | | | | |
| 20-80 | 430 | 0.7167 | 30 | 1.00 | 38 | 61.3 |
| 81-200 | 100 | 0.1667 | 45 | 1.25 | 12 | 19.4 |
| 201-300 | 50 | 0.0833 | 60 | 1.50 | 8 | 12.9 |
| 301-400 | 20 | 0.0333 | 75 | 2.00 | 4 | 6.5 |
| Using Equation 2-15, $n$ is equal to 60.0. Applying Equation 2-16 to each stratum yields a total of 62 samples after rounding up to the next integer. | | | | | | |

variability in the last three strata, a total of 59.3 or 60 samples are needed to meet the same design goal. So while more samples are taken in every stratum, proportionally fewer samples are needed in the smaller site size group. For example, using proportional allocation, more than 70 percent of the samples are taken in the 20- to 80-acre site size stratum, whereas approximately 57 percent of the samples are taken in the same stratum using the Neyman allocation.

Finally, introducing sample cost variation will also affect sample allocation. In the last scenario it was assumed that it is twice as expensive to evaluate a harvest site from the largest size stratum than to evaluate a harvest site from the smallest size stratum. In this example, roughly the same total number of samples are needed to meet the design goal, yet more samples are taken in the smaller size stratum.

### *2.3.3. Cluster Sampling*

Cluster sampling is commonly used when there is a choice between the size of the sampling unit (e.g., skid trail versus harvest site). In general, it is cheaper to sample larger units than smaller units, but the results tend to be less accurate (Snedecor and Cochran, 1980). Thus, if there is not a unit sampling cost advantage to cluster sampling, it is probably better to use simple random sampling. To decide whether to perform cluster sampling, it will probably be necessary to perform a special investigation to quantify sampling errors and costs using the two approaches.

Perhaps the best approach to explaining the difference between simple random sampling and cluster sampling is to consider an example set of results. In this example, the investigator did an evaluation to determine whether harvest sites had adequate SMAs. Since the state had timber harvesting activities across the state, the investigator elected to inspect 10 harvest sites along each randomly selected river. Table 2-7 presents the number of harvest sites along each river that had the recommended BMPs. The overall mean is 5.6; a little more than one-half of the sites have implemented the recommended BMPs. However, note that since the population unit corresponds to the 10 sites collectively, there are only 30 samples and the standard error for the proportion of sites using recommended BMPs is 0.035. Had the investigator incorrectly calculated the standard error using the random sampling equations, he or she would have computed 0.0287, nearly a 20 percent error.

Since the standard error from the cluster sampling example is 0.035, it is possible to estimate the corresponding simple random sample size to obtain the same precision using

$$n = \frac{pq}{s(p)^2} = \frac{(0.56)(0.44)}{0.035^2} = 201 \qquad \textbf{(2-17)}$$

Is collecting 300 samples using a cluster sampling approach cheaper than collecting about 200 simple random samples? If so, cluster sampling should be used; otherwise, simple random sampling should be used.

### *2.3.4. Systematic Sampling*

It might be necessary to obtain an estimate of the proportion of harvest sites where cable yarding was implemented using site inspections. Assuming a record of harvest sites (where cable yarding was specified in the timber sale contract or administration file) is available in a sequence unrelated to the manner in which this BMP would be implemented (e.g., in alphabetical order by

Table 2-7. Number of harvest sites (out of 10) implementing recommended BMPs.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 9 | 5 | 7 | 6 | 4 | 6 | 3 | 5 | 5 |
| 5 | 7 | 7 | 4 | 7 | 5 | 3 | 8 | 4 | 6 |
| 8 | 4 | 7 | 4 | 5 | 3 | 3 | 9 | 9 | 7 |

Grand Total = 168
$\overline{x} = 5.6$ $p = 5.6/10 = 0.560$
$s = 1.923$ $s = 1.923/10 = 0.1923$

Standard error using cluster sampling: $s(p)=0.1923/(30)^{0.5}=0.035$
Standard error if simple random sampling assumption had been incorrectly used:
$s(p)=((0.56)(1-0.56)/300)^{0.5}=0.0287$

the operator's name), a systematic sample can be obtained by selecting a random number $r$ between 1 and $n$, where $n$ is the number required in the sample (Casley and Lury, 1982). The sampling units are then $r$, $r + (N/n)$, $r + (2N/n)$, ..., $r + (n-1)(N/n)$, where $N$ is total number of available records.

If the population units are in random order (e.g., no trends, no natural strata, uncorrelated), systematic sampling is, on average, equivalent to simple random sampling.

Once the sampling units (in this case, specific harvest sites) have been selected, site inspections can be made to assess the extent of compliance with cable yarding standards and specifications.

# CHAPTER 3. METHODS FOR EVALUATING DATA

## 3.1. INTRODUCTION

Once data have been collected, it is necessary to statistically summarize and analyze the data. EPA recommends that the data analysis methods be selected before collecting the first sample. Many statistical methods have been computerized in easy-to-use software that is available for use on personal computers. Inclusion or exclusion in this section does not imply an endorsement or lack thereof by the U.S. Environmental Protection Agency. Commercial-off-the-shelf software that covers a wide range of statistical and graphical support includes SAS, Statistica, Statgraphics, Systat, Data Desk (Macintosh only), BMDP, and JMP. Numerous spreadsheets, database management packages, and other graphics software can also be used to perform many of the needed analyses. In addition, the following programs, written specifically for environmental analyses, are also available:

SCOUT: A Data Analysis Program, EPA, NTIS Order Number PB93-505303.

WQHYDRO (WATER QUALITY/HYDROLOGY GRAPHICS/ANALYSIS SYSTEM), Eric R. Aroner, Environmental Engineer, P.O. Box 18149, Portland, OR 97218.

WQSTAT, Jim C. Loftis, Department of Chemical and Bioresource Engineering, Colorado State University, Fort Collins, CO 80524.

Computing the proportion of sites implementing a certain BMP or the average number of acres that are under a certain BMP follows directly from the equations presented in Section 2.3 and is not repeated. The remainder of this section is focused on evaluating changes in BMP implementation. The methods provided in this section provide only a cursory overview of the type of analyses that might be of interest. For a more thorough discussion on these methods, the reader is referred to Gilbert (1987), Snedecor and Cochran (1980), and Helsel and Hirsch (1995). Typically the data collected for evaluating changes will typically come as two or more sets of random samples. In this case, the analyst will test for a shift or step change.

Depending on the objective, it is appropriate to select a one- or two-sided test. For example, if the analyst knows that BMP implementation will only go up as a result of an operator education program, a one-sided test could be formulated. Alternatively, if the analyst does not know whether implementation will go up or down, a two-sided test is necessary. To simply compare two random samples to decide whether they are significantly different, a two-sided test is used. Typical null hypotheses ($H_o$) and alternative hypotheses ($H_a$) for one- and two-sided tests are provided below:

<u>One-sided test</u>

$H_o$: BMP Implementation (Post education) $\leq$ BMP Implementation (Pre education)

$H_a$: BMP Implementation (Post education)
$>$ BMP Implementation (Pre education)

Two-sided test
$H_o$: BMP Implementation (Post education)
$=$ BMP Implementation (Pre education)

$H_a$: BMP Implementation (Post education)
$\neq$ BMP Implementation (Pre education)

Selecting a one-sided test instead of a two-sided test results in an increased power for the same significance level (Winer, 1971). That is, if the conditions are appropriate, a corresponding one-sided test is more desirable than a two-sided test given the same $\alpha$ and sample size. The manager and analyst should take great care in choosing one- or two-sided tests.

### 3.2. COMPARING THE MEANS FROM TWO INDEPENDENT RANDOM SAMPLES

The Student's $t$ test for two samples and the Mann-Whitney test are the most appropriate tests for these types of data. Assuming the data meet the assumptions of the $t$ test, the two-sample $t$ statistic with $n_1+n_2$-2 degrees of freedom is (Remington and Schork, 1970)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (3\text{-}1)$$

where $n_1$ and $n_2$ are the sample sizes of the first and second data sets, respectively, and $\bar{x}_1$ and $\bar{x}_2$ are the estimated means from the first and second data sets, respectively. The pooled standard deviation, $s_p$, is defined by

| Tests for Two Independent Random Samples | |
|---|---|
| **Test*** | **Key Assumptions** |
| Two-sample $t$ | • Both data sets must be normally distributed<br>• Data sets should have equal variances† |
| Mann-Whitney | • None |

* The standard forms of these tests require independent random samples.
† The variance homogeneity assumption can be relaxed.

$$s_p = \left[\frac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1+n_2-2}\right]^{0.5} \qquad (3\text{-}2)$$

where $s_1^2$ and $s_2^2$ correspond to the estimated variances of the first and second data sets, respectively. The difference quantity ($\Delta_o$) can be any value, but here it is set to zero. $\Delta_o$ can be set to a non-zero value to test whether the difference between the two data sets is greater than a selected value. If the variances are not equal, Snedecor and Cochran (1980) can be used as a source for methods for computing the $t$ statistic. In a two-sided test, the value from Equation 2-18 is compared to the $t$ value from Table A2 with $\alpha/2$ and $n_1+n_2$-2 degrees of freedom.

The Mann-Whitney test can also be used to compare two independent random samples. This test is very flexible since there are no assumptions about the distribution of either sample or whether the distributions have to be the same (Helsel and Hirsch, 1995). Wilcoxon (1945) first introduced this test for equal-sized samples. Mann and Whitney (1947) modified the original Wilcoxon's test to apply it to different sample sizes. Here, it is determined whether one data set tends to have larger observations than the other.

If the distributions of the two samples are similar except for location (i.e., similar spread and skew), $H_a$ can be refined to imply that the median concentration from one sample is "greater than," "less than," or "not equal to" the median concentration from the second sample. To achieve this greater detail in $H_a$, transformations such as logs can be used.

Tables of Mann-Whitney test statistics (e.g., Conover, 1980) can be consulted to determine whether to reject $H_o$ for small sample sizes. If $n_1$ and $n_2$ are greater than or equal to 10 observations, the test statistic can be computed from the following equation (Conover, 1980):

$$T_1 = \frac{T - n_1 \frac{n+1}{2}}{\sqrt{\frac{n_1 n_2}{n(n-1)} \sum_{i=1}^{n} R_i^2 - \frac{n_1 n_2 (n+1)^2}{4(n-1)}}} \quad \textbf{(3-3)}$$

where

$n_1$ = number of observations in sample with fewer observations,
$n_2$ = number of observations in sample with more observations,
$n$ = $n_1 + n_2$,
$T$ = sum of ranks for sample with fewer observations, and
$R_i$ = rank for the $i$th ordered observation used in both samples.

$T_1$ is normally distributed and Table A1 can be used to determine the appropriate quantile. Helsel and Hirsch (1995) and USEPA (1996) provide detailed examples for both of these tests.

## 3.3. Comparing the Proportions from Two Independent Samples

Consider the example in which the proportion of waterbars that effectively divert water from the skid trail has been estimated during two time periods to be $p_1$ and $p_2$ using sample sizes of $n_1$ and $n_2$, respectively. Assuming a normal approximation is valid, the test statistic under a null hypothesis of equivalent proportions (no change) is

$$\frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \textbf{(3-4)}$$

where $p$ is a pooled estimate of proportion and is equal to $(x_1+x_2)/(n_1+n_2)$ and $x_1$ and $x_2$ are the number of successes during the two time periods. An estimator for the difference in proportions is simply $p_1$-$p_2$.

In an earlier example, it was determined that 129 observations in each sample were needed to detect a difference in proportions of 0.20 with a two-sided test, $\alpha$ equal to 0.05, $1$-$\beta$ equal to 0.90. Assuming that 130 samples were taken and $p_1$ and $p_2$ were estimated from the data as 0.6 and 0.4, the test statistic would be estimated as

$$\frac{0.6 - 0.4}{\sqrt{0.5(0.5)\left(\frac{1}{130} + \frac{1}{130}\right)}} = 3.22 \quad \textbf{(3-5)}$$

Comparing this value to the $t$ value from Table A2 ($\alpha/2 = 0.025$, df=258) of 1.96, $H_o$ is rejected.

### 3.4. Comparing More Than Two Independent Random Samples

The analysis of variance (ANOVA) and Kruskal-Wallis are extensions of the two-sample $t$ and Mann-Whitney tests, respectively, and can be used for analyzing more than two independent random samples when the data are continuous (e.g., average SMA width). Unlike the $t$ test described earlier, the ANOVA can have more than one factor or explanatory variable. The Kruskal-Wallis test accommodates only one factor, whereas the Friedman test can be used for two factors. In addition to applying one of the above tests to determine if one of the samples is significantly different from the others, it is also necessary to perform postevaluations to determine which of the samples is different. This section recommends Tukey's method to analyze the raw or rank-transformed data only if one of the previous tests (ANOVA, rank-transformed ANOVA, Kruskal-Wallis, Friedman) indicates a significant difference between groups. Tukey's method can be used for equal or unequal sample sizes (Helsel and Hirsch, 1995). The reader is cautioned, when performing an ANOVA using standard software, to be sure that the ANOVA test used matches the data. USEPA (1996) provides a more detailed discussion on comparing more than two independent random samples.

### 3.5. Comparing Categorical Data

In comparing categorical data it is important to distinguish between whether the categories are nominal (e.g., land ownership, county location, type of BMP) or ordinal (e.g., BMP implementation rankings, low-medium-high scales).

The starting point for all evaluations is the development of a contingency table. In Table 3-1, the preference of three BMPs is compared to harvest site type in a contingency table. In this case both categorical variables are nominal. In this example, 45 of the 102 observations on federal lands used $BMP_1$. There were a total of 174 observations.

To test for independence, the sum of the squared differences between the expected ($E_{ij}$) and observed ($O_{ij}$) count summed over all cells is computed as (Helsel and Hirsch, 1995)

$$\chi_{ct} = \sum_{i=1}^{m} \sum_{j=1}^{k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \qquad \textbf{(3-6)}$$

where $E_{ij}$ is equal to $A_i C_j / N$. $\chi_{ct}$ is compared to the $1\text{-}\alpha$ quantile of the $\chi^2$ distribution with $(m\text{-}1)(k\text{-}1)$ degrees of freedom (see Table A3).

In the example presented in Table 3-1, the symbols listed in the parentheses correspond to the above equation. Note that $k$ corresponds to the three types of BMPs and $m$ corresponds to the three different types of harvest site. Table 3-2 shows computed values of $E_{ij}$ and $(O_{ij}\text{-}E_{ij})^2/E_{ij}$ in parentheses for the example data. $\chi_{ct}$ is equal to 14.60.

Table 3-1. Contingency table of harvest site type and implemented BMP.

| Harvest Site Type | $BMP_1$ | $BMP_2$ | $BMP_3$ | Row Total, $A_i$ |
|---|---|---|---|---|
| Private | 10 ($O_{11}$) | 30 ($O_{12}$) | 17 ($O_{13}$) | 57 ($A_1$) |
| Federal | 45 ($O_{21}$) | 32 ($O_{22}$) | 25 ($O_{23}$) | 102 ($A_2$) |
| State | 8 ($O_{31}$) | 3 ($O_{32}$) | 4 ($O_{33}$) | 15 ($A_3$) |
| Column Total, $C_j$ | 63 ($C_1$) | 65 ($C_2$) | 46 ($C_3$) | 174 (N) |

Key to Symbols:
$O_{ij}$ = number of observations for the *i*th harvest site and *j*th BMP type
$A_i$ = row total for the *i*th harvest site type (total number of observations for a given harvest site type)
$C_j$ = column total for the *j*th BMP type (total number of observations for a given BMP type)
N = total number of observations

From Table A3, the 0.95 quantile of the $\chi^2$ distribution with 4 degrees of freedom is 9.488. $H_0$ is rejected; the selection of BMP is not random among the different harvest site types. The largest values in the parentheses in Table 3-2 give an idea as to which combinations of harvest site type and BMP are noteworthy. In this example, it appears that $BMP_2$ is preferred to $BMP_1$ in comparison to federal and state harvest sites.

Now consider that in addition to evaluating information regarding the harvest site and BMP type, we also recorded a value from 1 to 5 indicating how well the BMP was installed and maintained, with 5 indicating the best results. In this case, the BMP implementation rating is ordinal. Using the same notation as before, the average rank of observations in row $x$, $R_x$, is equal to (Helsel and Hirsch, 1995)

$$R_x = \sum_{i=1}^{x-1} A_i + (A_x+1)/2 \qquad (3\text{-}7)$$

where $A_i$ corresponds to the row total. The average rank of observations in column $j$, $D_j$, is equal to

$$D_j = \frac{\sum_{i=1}^{m} O_{ij} R_i}{C_j} \qquad (3\text{-}8)$$

where $C_j$ corresponds to the column total. The Kruskal-Wallis test statistic is then computed as

$$K = (N-1)\,\frac{\sum_{j=1}^{k} C_j D_j^2 - N\left[\frac{N+1}{N}\right]^2}{\sum_{i=1}^{m} A_i R_i^2 - N\left[\frac{N+1}{N}\right]^2} \qquad (3\text{-}9)$$

where $K$ is compared to the $\chi^2$ distribution with *k-1* degrees of freedom. This is the most general form of the Kruskal-Wallis test since it is a comparison of distribution shifts rather than shifts in the median (Helsel and Hirsch, 1995).

Table 3-2. Contingency table of expected harvest site type and implemented BMP. (Values in parentheses correspond to $(O_{ij}-E_{ij})^2/E_{ij}$.)

| Harvest Site Type | $BMP_1$ | $BMP_2$ | $BMP_3$ | Row Total |
|---|---|---|---|---|
| Private | 20.64 (5.48) | 21.29 (3.56) | 15.07 (0.25) | 57 |
| Federal | 36.93 (1.76) | 38.10 (0.98) | 26.97 (0.14) | 102 |
| State | 5.43 (1.22) | 5.60 (1.21) | 3.97 (0.00) | 15 |
| Column Total | 63 | 65 | 46 | 174 |

Table 3-3 is a continuation of the previous example indicating the BMP implementation rating for each BMP type. For example, 29 of the 70 observations that were given a rating of 4 are associated with $BMP_2$. The terms inside the parentheses of Table 3-3 correspond to the terms used in Equations 3-7 to 3-9. Note that $k$ corresponds to the three types of BMPs and $m$ corresponds to the five different levels of BMP implementation. Using Equation 3-9 for the data in Table 3-3, $K$ is equal to 14.86. Comparing this value to 5.991 obtained from Table A3, there is a significant difference in the quality of implementation between the three BMPs.

The last type of categorical data evaluation considered in this chapter is that in which both variables are ordinal. The Kendall $\tau_b$ for tied data can be used for this analysis. The statistic $\tau_b$ is calculated as (Helsel and Hirsch, 1995)

$$\tau_b = \frac{S}{\frac{1}{2}\sqrt{(N^2-SS_a)(N^2-SS_b)}} \qquad \textbf{(3-10)}$$

where $S$, $SS_a$, and $SS_c$ are computed as

$$S = \sum_{all\,xy}\left[\sum_{i>x}\sum_{j>y} O_{xy}O_{ij} - \sum_{i<x}\sum_{j<y} O_{xy}O_{ij}\right] \qquad \textbf{(3-11)}$$

$$SS_a = \sum_{i=1}^{m} A_i^2 \qquad \textbf{(3-12)}$$

$$SS_c = \sum_{j=1}^{k} C_j^2 \qquad \textbf{(3-13)}$$

To determine whether $\tau_b$ is significant, $S$ is modified to a normal statistic, using

$$Z_S = \begin{cases} \dfrac{S-1}{\sigma_S} & if\ S>0 \\ \dfrac{S+1}{\sigma_S} & if\ S<0 \end{cases} \qquad \textbf{(3-14)}$$

Table 3-3. Contingency table of implemented BMP and rating of installation and maintenance.

| BMP Implementation Rating | $BMP_1$ | $BMP_2$ | $BMP_3$ | Row Total, $A_i$ |
|---|---|---|---|---|
| 1 | 1 ($O_{11}$) | 2 ($O_{12}$) | 2 ($O_{13}$) | 5 ($A_1$) |
| 2 | 7 ($O_{21}$) | 3 ($O_{22}$) | 5 ($O_{23}$) | 15 ($A_2$) |
| 3 | 15 ($O_{31}$) | 16 ($O_{32}$) | 26 ($O_{33}$) | 57 ($A_3$) |
| 4 | 32 ($O_{41}$) | 29 ($O_{42}$) | 9 ($O_{43}$) | 70 ($A_4$) |
| 5 | 8 ($O_{51}$) | 15 ($O_{52}$) | 4 ($O_{53}$) | 27 ($A_5$) |
| Column Total, $C_j$ | 63 ($C_1$) | 65 ($C_2$) | 46 ($C_3$) | 174 (N) |

Key to Symbols:
$O_{ij}$ = number of observations for the *i*th BMP implementation rating and *j*th BMP type
$A_i$ = row total for the *i*th BMP implementation rating (total number of observations for a given BMP implementation rating)
$C_j$ = column total for the *j*th BMP type (total number of observations for a given BMP type)
N = total number of observations

where

$$\sigma_s = \sqrt{\frac{N^3}{9}\left(1-\sum_{i=1}^{m} a_i^3\right)\left(1-\sum_{j=1}^{k} c_j^3\right)} \quad \textbf{(3-15)}$$

where $Z_S$ is zero if $S$ is zero. The values of $a_i$ and $c_i$ are compute as $A_i/N$ and $C_i/N$, respectively.

Table 3-4 presents the BMP implementation ratings that were taken in three separate years. For example, 15 of the 57 observations that were given a rating of 3 are associated with Year 2. Using Equations 3-11 and 3-15, $S$ and $\sigma_s$ are equal to 2,509 and 679.75, respectively. Therefore, $Z_s$ is equal to (2509-1)/679.75 or 3.69. Comparing this value to a value of 1.96, obtained from Table A1 ($\alpha/2=0.025$), indicates that BMP implementation is improving with time.

Table 3-4. Contingency table of implemented BMP and sample year.

| BMP Implementation Rating | Year 1 | Year 2 | Year 3 | Row Total, $A_i$ | $a_i$ |
|---|---|---|---|---|---|
| 1 | 2 ($O_{11}$) | 1 ($O_{12}$) | 2 ($O_{13}$) | 5 ($A_1$) | 0.029 |
| 2 | 5 ($O_{21}$) | 7 ($O_{22}$) | 3 ($O_{23}$) | 15 ($A_2$) | 0.086 |
| 3 | 26 ($O_{31}$) | 15 ($O_{32}$) | 16 ($O_{33}$) | 57 ($A_3$) | 0.328 |
| 4 | 9 ($O_{41}$) | 32 ($O_{42}$) | 29 ($O_{43}$) | 70 ($A_4$) | 0.402 |
| 5 | 4 ($O_{51}$) | 8 ($O_{52}$) | 15 ($O_{53}$) | 27 ($A_5$) | 0.155 |
| Column Total, $C_j$ | 46 ($C_1$) | 63 ($C_2$) | 65 ($C_3$) | 174 (N) | |
| $c_j$ | 0.264 | 0.362 | 0.374 | | |

Key to Symbols:
$O_{ij}$ = number of observations for the *i*th BMP implementation rating and *j*th year
$A_i$ = row total for the *i*th BMP implementation rating (total number of observations for a given harvest type)
$C_j$ = column total for the *j*th BMP type (total number of observations for a given year)
N = total number of observations
$a_i$ = $A_i$/N
$c_j$ = $C_j$/N

# CHAPTER 4. CONDUCTING THE EVALUATION

## 4.1 INTRODUCTION

This chapter addresses the process of determining whether forestry MMs or BMPs are being implemented and whether they are being implemented according to approved standards or specifications. Guidance is provided on what should be measured to assess MM and BMP implementation, as well as methods for collecting the information, including physical site evaluations, mail- and/or telephone-based surveys, personal interviews, and aerial reconnaissance and photography. Designing survey instruments to avoid error and rating MM and BMP implementation are also discussed.

Evaluation methods are separated into two types: Expert evaluations and self-evaluations. Expert evaluations are those in which actual field investigations are conducted by trained personnel to gather information on MM or BMP implementation. Self-evaluations are those in which answers to a predesigned questionnaire or survey are provided by harvesters and/or landowners associated with the survey site. Self-evaluations might also include examination of materials related to a harvest, such as harvest plans and records of violations of forestry regulations. Extreme caution should be exercised when using data from self-evaluations as the basis for assessing MM or BMP implementation since they are not typically reliable for this purpose. Each of these evaluation methods has advantages and disadvantages that should be considered before to deciding which one to use or in what combination to use them. Aerial reconnaissance and photography can be used to support either evaluation method.

Self-evaluations are useful for collecting information on landowner or harvester awareness of MMs or BMPs, dates of harvest, harvest site conditions, which MMs or BMPs were implemented, and whether the assistance of a professional forester was used. However, the type of or level of detail of information that can be obtained from self-evaluations might be inadequate to satisfy the objectives of a MM/BMP compliance survey. If this is the case, expert evaluations might be called for. Expert evaluations are necessary if information on MM/BMP implementation that is more detailed or more reliable than that that can be obtained with self-evaluations is required, such as an objective assessment of the adequacy of MM/BMP implementation, the degree to which site-specific factors (e.g., slope, soil type, or presence of a water body) influenced MM/BMP implementation, or the need for changes in standards and specifications for MM/BMP implementation. Sections 4.3 and 4.4 discuss expert evaluations and self-evaluations, respectively, in more detail.

Expert evaluations of implementation of forestry MMs or BMPs generally occur after the harvest has occurred (see ***Example***), and direct observation of the adequacy of implementation of many BMPs (e.g., preharvest planning, pesticide applications, or construction of temporary roads) might not be possible. However, evidence of proper BMP implementation is often present at harvest sites. For instance, evidence of excessive skidding in SMAs, vegetation kills due to pesticide use in SMAs, and poorly restored stream banks and stream beds where temporary stream crossings were located is often observable during site evaluations. Supplemental information on aspects of harvest operations that cannot be observed directly during site evaluations might also be obtained from self-evaluations.

***Example ...*** Timing of site evaluations.

> *U.S. Forest Service, Southwest Region*: After logging to within approximately one year after the site is harvested (USDA, 1992).
> *South Carolina*: One year or less than after the site is harvested (Adams, 1994).
> *Florida*: Two years or less after the site is harvested (Vowell and Gilpin, 1994).

Aerial reconnaissance and photography is another means available for collecting information on harvests, though many of the MMs/BMPs employed for forestry are difficult if not impossible to identify on aerial photographs. For this reason, aerial reconnaissance and photography are most useful for identifying potential survey sites and monitoring harvest site regeneration, forest conditions, and some water quality conditions (e.g., sediment runoff, algal blooms). Aerial reconnaissance and photography are discussed in more detail in Section 4.5.

The general types of information obtainable with self-evaluations are listed in Table 4-1. Regardless of the approach(es) used, proper and thorough preparation for the evaluation is the key to success.

## 4.2 Choice of Variables

Once the objectives of a BMP implementation or compliance survey have been clearly defined, the most important factor in the assessment of MM or BMP implementation is the determination of which variable(s) to measure. A good variable provides a direct measure of how well a BMP was implemented. Individual variables should provide measures of different factors related to BMP implementation. The best variables are those which are measures of the adequacy of MM or BMP implementation and are based on quantifiable expressions of conformance with state standards and specifications. As the variables used become less directly related to actual MM or BMP implementation, their accuracy as measures of BMP implementation decreases.

Examples of useful variables include width of streamside management areas, slope of landing areas, and size of culverts, all of which would be expressed in terms of conformance with applicable state standards and specifications. Less useful variables measure factors that are related to BMP implementation but do not necessarily provide an accurate measure of their implementation. Examples of such variables include the number of miles of forest roads constructed and the number of preharvest plans submitted to the state forestry agency, department, or division (hereafter referred to as the state forestry authority). Although these variables relate to MM and BMP

Table 4-1. General types of information obtainable with self-evaluations and expert evaluations.

| Information Obtainable from Self-Evaluations |
| --- |
| *Harvest applications and management plans associated with the harvest (e.g., preharvest, road, fire, forest chemical) might be available for review prior to site evaluations and can provide much background information, such as:*<br>• Type of ownership (private industrial, private nonindustrial, federal, other public)<br>• Total acreage under management<br>• Acres/board feet harvested<br>• Surface water body types on harvest site<br>• Soil type<br>• Ecological characterization of harvested area (e.g., habitat type, significant wildlife)<br>• Presence of critical wildlife habitat<br>• Species harvested<br>• Harvest/management history of the harvested site<br>• Use of cable yarding or ground skidding<br>• Chemicals (e.g., pesticides, fertilizers) applied<br>• Dates of plan preparation and revisions<br>• Locations of roads and road structures, SMAs, loading areas, etc.<br>• MMs and BMPs applied during harvest<br>• Map<br><br>*Conversations with harvesters and landowners can be used to verify information obtained from applications or can yield supplemental information, such as:*<br>• Dates of harvest<br>• Ambient conditions during applications<br>• Variations from harvest plan<br>• Problems encountered during harvest<br>• Types of equipment used during harvest and in SMAs<br>• Timing, location, and rate of chemical applications |
| **Information Requiring Expert Evaluations** |
| *Expert evaluations are necessary to verify information obtained from self-evaluations and records and to assess the actual adequacy of MM and BMP implementation. Expert evaluations are necessary to:*<br>• Assess design adequacy<br>• Assess installation adequacy<br>• Assess the appropriateness of operation methods and overall management<br>• Confirm information obtained from self-evaluations |

implementation, they provide no real information on whether the MMs and BMPs are actually being implemented or whether they are being implemented properly.

Variables generally will not directly relate to MM implementation since most forestry MMs are combinations of several BMPs. Measures of MM implementation, therefore, usually will be based on separate assessments of two or more BMPs, and the implementation of each BMP will be based on a unique set of variables. Some examples of BMPs related to EPA's Road Construction and Reconstruction Management Measure, variables for assessing compliance with the BMPs, and related standards and specifications that might be required by state forestry authorities are presented in Figure 4.1. Because harvesters choose to implement or not implement MMs/BMPs based on site-specific conditions, it is also appropriate to apply varying weights to the variables chosen to assess MM/BMP implementation to correspond to site-specific conditions. For example, variables related to slope factors might be de-emphasized—and other, more applicable variables emphasized more—on relatively flat harvest sites. Similarly, on a site with a water body, variables related to SMAs, sediment runoff, and chemical deposition (pesticide use, fertilizer use) might be emphasized over other variables to arrive at a site-specific rating of the adequacy of MM/BMP implementation.

The purpose for which the information collected during an MM or BMP implementation survey will be used is another important consideration when selecting variables. An implementation survey can serve many purposes beyond the primary purpose of assessing MM and BMP implementation. For instance, for its 1993 BMP compliance survey, the South Carolina Forestry Commission selected variables that enabled it to assess compliance with each of five categories of BMPs and overall compliance with BMPs. In addition, the Commission analyzed the effect of each of 16 additional variables on BMP compliance (see ***Example***). The purpose of the survey was not only to assess BMP implementation, but also to assess the relationship of various conditions to the level of BMP compliance (Adams, 1994).

Table 4-2 provides examples of useful and less useful variables for the assessment of implementation of the forestry MMs developed by EPA (USEPA, 1993a). The variables listed in the table are only examples, and local or regional conditions ultimately dictate which variables should be used.

## 4.3 Expert Evaluations

### 4.3.1 Site Evaluations

Expert evaluations are the best way to collect reliable information on MM and BMP implementation. They involve a person or team of people visiting individual harvest sites and speaking with harvest operators and/or landowners to obtain information on MM and BMP implementation. For many MMs, assessing and verifying compliance requires a site visit and evaluation. The

**Management Measure for Road Construction and Reconstruction**

**(1) Follow preharvest planning (as described in the Preharvest Planning Management Measure) when constructing or reconstructing the roadway.**
**(2) Follow designs planned under the Preharvest Planning Management Measure for road surfacing and shaping.**
**(3) Install road drainage structures according to designs planned under the Preharvest Planning Management Measure and regional storm return period and installation specifications. Match these drainage structures with terrain features and with road surface and prism designs.**
**(4) Guard against the production of sediment when installing stream crossings.**
**(5) Protect surface waters from slash and debris material from roadway clearing.**
**(6) Use straw bales, silt fences, mulching, or other favorable practices on disturbed soils on unstable cuts, fills, etc.**
**(7) Avoid constructing new roads in SMAs to the extent practicable.**

*Related BMPs, measurement variables, and standards and specifications*:

| Management Measure Practice | Potential Measurement Variables | Example Related Standards and Specifications |
|---|---|---|
| • Preplan skid trail and landing locations on stable soils and avoid steep gradients and areas that are landslide-prone or erosion-prone, or have poor drainage. | • Soil type or stability along skid trails and at landings.<br>• Gradients along skid trails and at landings. | • Minimum soil stability for skid trails and landings.<br>• Maximum slope for skid trails and landings. |
| • In moderately sloping terrain, plan for road grades of less than 10%, with an optimal grade between 3% and 5%. Vary road grades frequently to reduce culvert and road drainage ditch flows, road surface erosion, and concentrated culvert discharges. | • Categorization of terrain as flat, moderate, steep.<br>• Road grade estimated over sections of road.<br>• Steep terrain: Average distance between changes in grade. | • Maximum road grade for a given terrain slope.<br>• Maximum distance between changes in grade on steep terrain.<br>• Minimum/maximum distance between drainage features for a given terrain slope. |
| • Design roads and skid trails to follow the natural topography and contour, minimizing alteration of natural features. | • Natural slope of surrounding terrain.<br>• Slope of skid trails and at landings. | • Maximum slope of skid trails for a given terrain slope. |
| • Design cut-and-fill slopes to be at stable angles, or less than the normal angle of repose, to minimize erosion and slope failure potential. | • Angle of cut-and-fill slopes.<br>• Stability of soil type where cut-and-fill slopes have been installed. | • Maximum angle for cut-and-fill slopes. |

Figure 4-1. Potential variables and examples of implementation standards and specifications that might be useful for evaluating compliance with the Road Construction and Reconstruction Management Measure.

Table 4-2. Examples of variables related to management measure implementation.

| Management Measure | Useful Variables | Less Useful Variables | Appropriate Sampling Unit |
|---|---|---|---|
| Preharvest Planning | · Agreement between preharvest plan and harvest operation<br>· Inclusion of all required elements in preharvest plan | · Number of preharvest plans developed/approved | · Harvest operation<br>· Preharvest plan |
| Streamside Management Areas | · Width of SMAs<br>· Leave trees in SMAs meet minimum requirements | · Presence of water body on harvest site<br>· Number of stream crossings in SMA | · 100-ft stretch of SMA |
| Road Construction/ Reconstruction | · Compaction of fill materials adequate to prevent erosion<br>· Culverts cross streams at right angles | · Miles of road constructed<br>· Number of stream crossings installed | · Fill areas along forest roads<br>· Stream crossings |
| Road Management | · Culverts free of obstructions<br>· Temporary stream crossings removed | · Completion of road inspections<br>· Number of temporary stream crossings removed | · Culverts<br>· Forest road stream crossings |
| Timber Harvesting | · Proper slope at landings<br>· Water bodies free of slash materials | · Acres harvested<br>· Number of cable yarding operations | · Landings<br>· 100 yd of stream adjacent to harvest site |
| Site Preparation and Forest Regeneration | · Adequate distribution of seedlings on prepared sites<br>· Nonmechanical site preparation used in SMAs | · Method of site preparation<br>· Acres revegetated | · 100-$yd^2$ plots<br>· 100 yd of SMA |
| Fire Management | · Fire lines constructed to minimize erosion<br>· Intense burning not conducted on steep slopes with high erosion potential | · Acres burned<br>· Size of individual burn areas | · 100 yd of fire line<br>· Burned areas |
| Regeneration of Disturbed Areas | · Minimum requirements for seedlings per acre met<br>· Erosion-prone areas replanted | · Species planted<br>· Acres revegetated | · Steeply sloped areas<br>· 100-$yd^2$ plots |
| Forest Chemical Management | · Mixing and loading areas located away from surface waters<br>· Pesticides applied in accordance with EPA and/or state requirements | · Pounds of chemical applied<br>· Availability of spill contingency plan | · SMAs<br>· Chemical mixing and loading areas<br>· Harvest site |
| Wetlands Forest | · Any of above with respect to wetlands forest | · As above | · As above |

***Example ...*** Variables used by the South Carolina Forestry Commission during the 1993 BMP compliance survey. (Source: Adams, 1994)

> *Overall BMP compliance and compliance with each of five categories of BMPs were assessed:*
>
> - Road systems
> - Road stream crossings
> - Streamside management zones
> - Log decks
> - Harvesting operations
>
> and for overall BMP compliance.
>
> *Sixteen additional variables were analyzed to determine their relationship to BMP compliance:*
>
> - Presence of perennial streams
> - Terrain type
> - Percent slope
> - Use of a professional forester
> - Required compliance with BMPs
> - Physiographic region
> - Use of a sales contract
> - Percent of site impacted
> - Landowner category
> - Familiarity of landowner with BMPs
> - Logged under wet soil conditions
> - Rutting severity
> - Road construction applicability
> - Soil drainage class
> - Presence of jurisdictional wetlands
> - Harvest size

following should be considered before expert evaluations are conducted:

- *Obtaining permission from the landowner*. Without proper authorization to visit a site from a landowner, the relationship between landowners and the state forestry authority, and any future regulatory or compliance action could be jeopardized.

- *The type(s) of expertise needed to assess proper implementation.* For some MMs, a team of trained personnel might be required at a site evaluation to determine whether MMs have been implemented properly.

- *The activities that should occur during a site evaluation.* This information is necessary for proper and complete preparation for the site visit, so that the evaluation can be completed in a single visit and at the proper time.

- *The method of rating the MMs/BMPs.* MM and BMP rating systems are discussed below.

- *Consistency among evaluation teams and among site evaluations.* Proper training and preparation of site evaluation team members are crucial to ensure accuracy and consistency.

- *The collection of information while at a site.* Information collection should be facilitated with preparation of data collection forms that include any necessary MM and BMP rating information needed by the evaluation team members.

- *The content and format of postevaluation discussions.* Site evaluation team members should bear in mind the value of postevaluation discussion among team members. Notes can be taken during the evaluation concerning any items that would benefit from group discussion.

Evaluators might consist of a single person suitably trained in silvicultural site evaluation to a group of professionals with various expertise. The composition of an evaluation team will depend on the types of MMs or BMPs being evaluated. Potential team members could include:

- Forester
- State forestry personnel
- Hydrologist
- Pesticide specialist
- Soil scientist
- Water quality expert

The composition of evaluation teams can vary depending on the purpose of the evaluation, available staff and other resources, and the geographic area being covered. All team members should be familiar with the required MMs/BMPs, and each team should have a member who has previously participated in a site evaluation. This will ensure familiarity with the technical aspects of the MMs/BMPs that will be rated during the evaluation and the site evaluation process.

Training may be necessary to bring all team members to the level of proficiency needed to conduct the site evaluations. State forestry personnel should be familiar with forestry regulations, state BMP standards and specifications, and proper BMP implementation, and therefore are generally well qualified to teach these topics to evaluation team members who are less familiar with them. This training should include identification of BMPs particularly critical to water quality protection, BMPs implemented poorly in previous years, analysis of erosion potential, and other aspects of BMP implementation that require professional judgement, as well as any standard methods for measurements to judge BMP implementation against state standards and specifications.

Alternatively, if only one or two individuals will be conducting site evaluations, their training in the various specialties, such as those listed above, necessary to evaluate the quality of MM/BMP implementation could be provided by a team of specialists who are familiar with forestry practices and nonpoint source pollution.

In the interest of consistency among the evaluations and among team members, it is advisable that one or more mock evaluations take place prior to visiting selected sample sites. These "practice sessions" provide team members with an opportunity to become familiar with MMs and BMPs as they should be implemented under different harvest site conditions, gain familiarity with the evaluation forms and the meanings of the terms and questions on them, and learn from other team members with different expertise. Mock evaluations are valuable for ensuring that all evaluators have a similar understanding of the intent of the questions, especially for questions whose responses involve a degree of subjectivity on the part of the evaluator.

Where site evaluation teams are composed of more than two or three people, it might be helpful to divide the various responsibilities for conducting the site evaluations among team members ahead of time to avoid confusion at the harvest site and to be certain that all tasks are completed but not duplicated. Having a spokesperson for the group who is responsible for communicating with the landowner or harvester—prior to the site evaluation, at the site evaluation if they are present, and afterward—might also be helpful. A state forestry representative is generally a good choice as spokesperson because he/she represents the state forestry authority. Newly formed evaluation teams might benefit most from a division of labor and selection of a team leader or team coordinator with experience with site evaluations who will be responsible for the quality of the site evaluations. Smaller teams might find that a division of responsibilities is not necessary, as might larger teams whose members have experience working together. If responsibilities are to be assigned, mock evaluations can be a good time to work out these details.

### 4.3.2 Rating Implementation of Management Measures and Best Management Practices

Many factors influence the implementation of MMs and BMPs, so it is sometimes necessary to use best professional judgment (BPJ) to rate their implementation and BPJ will almost always be necessary when rating the implementation of MMs or when rating overall BMP compliance at a harvest site. Site-specific factors such as soil type, slope, presence of a water body, and ground cover type affect the implementation of erosion and sediment control BMPs, for instance, and must be taken into account by evaluators when rating MM/BMP implementation. Implementation of MMs will often be based on implementation of more than one BMP, and this makes rating MM implementation similar to rating overall BMP implementation at a harvest site. Determining an overall rating involves grouping the ratings of implementation of individual BMPs into a single rating, which introduces more subjectivity than rating the implementation of individual BMPs based on standards and specifications. Choice of a rating system and rating terms, which are aspects of proper evaluation design, is therefore important in minimizing the level of subjectivity associated with overall BMP compliance and MM implementation ratings. When creating overall ratings, it is still important to record the detailed ratings of individual BMPs as supporting information.

Individual BMPs, overall BMP compliance, and MMs can be rated using a binary approach (e.g., pass/fail, compliant/ noncompliant, or yes/no) or on a scale with

***Example ...*** of a rating scale (Source: Rossman and Phillips, 1992). More examples are presented in Appendix B.

> Minnesota Division of Forestry uses this 5-choice rating scale for BMP implementation audits:
> 5 = Operation exceeds requirement of BMP
> 4 = Operation meets requirement of BMP
> 3 = Minor departure from BMP
> 2 = Major departure from BMP
> 1 = Gross neglect of BMP
>
> where:
>
> *Minor departure* is defined as "small in magnitude and localized," *major departure* is defined as "significant magnitude or where the BMPs were consistently neglected" and *gross neglect* is defined as "potential risk to water resources was significant and there was no evidence that any attempt has been made by the operator to apply the BMP."

more than two choices, such as 1 to 5 or 1 to 10 (where 1 is the worst) (see ***Example***). The simplest method of rating MM and BMP implementation is the use of a binary approach. Using a binary approach, either an entire site or individual MMs or BMPs are rated as being in compliance or not in compliance with respect to specified criteria. Scale systems can take the form of ratings from poor to excellent, inadequate to adequate, low to high, 1 to 3, 1 to 5, and so forth.

Whatever form of scale is used, the factors that would individually or collectively qualify a site, MM, or BMP for one of the ratings should be clearly stated. The more choices that are added to the scale, the smaller and smaller the difference between them becomes and each must therefore be defined more specifically and accurately. This is especially important if different teams or individuals rate separate sites. Consistency among the ratings then depends on each team or individual evaluator knowing precisely what the criteria for each rating option mean. Clear and precise explanations of the rating scale can also help avoid or reduce disagreements among team members. This applies equally to a binary approach. The factors, individually or collectively, that would cause a site, MM, or BMP to be rated as not being in compliance with design specifications should be clearly stated on the evaluation form or in supporting documentation.

Rating sites or MMs/BMPs on a scale requires a greater degree of analysis by the evaluation team than does using a binary approach. Each higher number represents a better level of MM/BMP implementation and/or effectiveness. In effect, a binary rating approach is a scale with two choices; a scale of low, medium, and high (compliance) is a scale with three choices. Use of a scale system with more than two rating choices can provide more information to program managers than a binary rating approach, and this factor must be weighed against the greater complexity involved in using one. For instance, a survey that uses a scale of 1 to 5 might result in one MM with a rating of 1, five with a rating of 2, six with a rating of 3, eight with a rating of 4, and five with a rating of 5. Precise criteria would have to be developed to be able to ensure consistency within and between survey teams in rating the MMs, but the information that only 1 MM was poorly implemented, 11 were below standards, 13 met or were above standards, and 5 were implemented very well might be more valuable than the information that 18 MMs were found to be in compliance with design specifications, which is the only information that would be obtained with a binary rating approach.

If a rating system with more than two ratings is used to collect data, the data can be analyzed either by using the original rating data or by first transforming the data into a binomial (i.e., two-choice rating) system. For instance, ratings of 1 through 5 could be reduced to two ratings by grouping the 1s, 2s, and 3s together into one group (e.g., inadequate) and the 4s and 5s into a separate group (e.g., adequate). If this approach is used, it is best to retain the rating data for the detailed information it contains and to reduce the data to a binomial system only for the purpose of statistical analysis. Chapter 3, Section 3.5, contains information on the analysis of categorical data.

### 4.3.3 Rating Terms

The choice of rating terms used on the evaluation forms is an important factor in ensuring consistency and reducing bias, and the terms used to describe and define the rating options should be as objective as possible. For a rating system with a large number of options, the meanings of each option should be clearly defined. It is best to avoid using terms such as "major" and "minor" when describing erosion or pollution effects or deviations from prescribed MM/BMP implementation criteria because they might have different connotations for different evaluation team members. It is easier for an evaluation team to agree on meaning if options are described in terms of measurable criteria and examples are provided to clarify the intended meaning. It is also best not to use terms that carry negative connotations. Evaluators are less likely to rate something as having a "major deviation" from an implementation criterion, even if justified, because of the negative connotation carried by the term. Rather than using such a term, observable conditions or effects of the quality of implementation should be listed and specific ratings (e.g., 1-5 or compliant/noncompliant for the criterion) should be associated with the conditions or effects. For example, instead of rating culvert installation as having a "major deficiency," a specific deficiency should be described and should have an associated rating ascribed to it (e.g., "Culvert as installed does not allow for fish passage = noncompliant").

Evaluation team members will often have to take specific notes on sites, MMs, or BMPs during the evaluation, either to justify the ratings they have ascribed to variables or for discussion with other team members after the survey. When recording notes about the sites, MMs, or BMPs, evaluation team members should be as specific as the criteria for the ratings. A rating recorded as "MM deviates highly from implementation criteria" is highly subjective and loses specific meaning when read by anyone other than the person who wrote the note. Notes should therefore be as objective and specific as possible.

An overall site rating is useful for summarizing information in reports; identifying the overall level of implementation with MMs/BMPs, indicating the likelihood that environmental protection is being achieved, identifying additional training or education needs; and conveying information to program managers who are often not familiar with MMs or BMPs. For the purposes of preserving the valuable information contained in the original ratings of sites, MMs, or BMPs, however, overall ratings should summarize, not replace the original data. Analysis of year-to-year variations in MM or BMP implementation, the factors involved in MM or BMP program implementation, and factors that could improve MM or BMP implementation and MM or BMP program success are possible only if the original, detailed site, MM, or BMP data are used.

Approaches commonly used for determining final BMP implementation ratings include calculating a percentage based on individual BMP ratings, consensus, compilation of aggregate scores by an objective party, voting, and voting only where consensus on a site or MM/BMP rating cannot be reached. Not all systems for arriving at final ratings are applicable to all circumstances.

### *4.3.4 Consistency Issues*

Consistency among evaluators and between evaluations is important. Consistency is likely to be best if only one or two evaluators conduct the site evaluations and the same persons conduct all of the evaluations. If, for statistical purposes, many sites (e.g., 100 or more) need to be evaluated, use of only one or two evaluators might also be the most efficient approach. In this case, a team of evaluators might be useful for revisiting a subsample of the sites evaluated by the one to two persons for quality control purposes. Evaluation teams can also be useful for training the one or two persons who will conduct the site evaluations in their specialties as they relate to MM/BMP implementation and nonpoint source pollution.

If teams of evaluators conduct the evaluations, consistency can be achieved by keeping the membership of the teams constant. Differences of opinion, which are likely to arise among team members, can be settled through discussions held during evaluations, and the experience of team members who have done past evaluations can help guide decisions. Pre-evaluation training sessions, such as the mock evaluations discussed above, will help ensure that the first few site evaluations are not "learning" experiences to such an extent that sites must be revisited to ensure that they receive the same level of scrutiny as sites evaluated later.

If different sites are visited by different teams of evaluators or if individual evaluators are assigned to different sites, it is especially important that consistency be established before the evaluations are conducted. For best results, discussions among evaluators should be held periodically during the evaluations to discuss any potential problems. For instance, evaluators could visit some sites together at the beginning of the evaluations to promote consistency in ratings, followed by site evaluations conducted by individual evaluators. Then, after a few site or MM evaluations, evaluators could gather again to discuss results and to share any knowledge gained to ensure continued consistency.

As mentioned above, consistency can be established during mock evaluations held before the actual evaluations begin. These mock evaluations are excellent opportunities for evaluators to discuss the meaning of terms on rating forms, differences between rating criteria, and differences of opinion about proper MM/BMP implementation. A member of the evaluation team should be able to represent the state's position on the definition of terms and clarify areas of confusion.

Descriptions of MMs and BMPs should be detailed enough to support any ratings given to individual features and to the MM or BMP overall. Sketching a diagram of the MM or BMP helps identify design problems, promotes careful evaluation of all features, and provides a record of the MM or BMP for future reference. A diagram is also valuable when discussing the MM or BMP with the landowner or identifying features in need of improvement or alteration. Landowners can also use a copy of the diagram and evaluation when discussing their operations with state forestry agents. Photographs of MM or BMP features are valuable reference material and should be used whenever an evaluator feels that a written description or a diagram could be inadequate. Photographs of what constitutes both good and poor MM or BMP implementation are valuable for explanatory and educational purposes; for example, for presentations to managers and the public.

### 4.3.5 Postevaluation Onsite Activities

It is important to complete all pertinent tasks as soon as possible after the completion of a site evaluation to avoid extra work later and to reduce the chances of introducing error attributable to memory error or confusion. All evaluation forms for each site should be filled out completely before leaving the site. Information not filled in at the beginning of the evaluation can be obtained from the landowner if necessary. Any questions that evaluators had about the MMs/BMPs during the evaluation can be discussed, and notes written during the evaluation can be shared and used to help clarify details of the evaluation process and ratings. The opportunity to revisit the site will still exist if there are points that cannot be agreed upon among evaluation team members.

Also, while the evaluation team is still on site, the landowner should be informed about what will follow; for instance, whether he/she will receive a copy of the report, when to expect it, what the results means, and his/her responsibility in light of the evaluation, if any. Immediately following the evaluation is also an excellent time to discuss the findings with the landowner if he/she was not present during the evaluation.

## 4.4 SELF-EVALUATIONS

### 4.4.1 Methods

Self-evaluations, while often not a reliable source of MM or BMP implementation data, can be used to augment data collected through expert evaluations or in place of expert evaluations where the latter cannot be conducted. In some cases, state forestry authority staff might have been involved directly with a harvest and will be a source of useful information even if an expert evaluation is not conducted. Self-evaluations are an appropriate survey method for obtaining background information from harvesters and landowners.

Mail, telephone, and mail with telephone follow-up are common self-evaluation methods. Mail and telephone surveys are useful for collecting general information, such as the location of harvest operations, species harvested, methods used, and dates of harvest. Also, harvest application or notification records can provide useful background information, including any special conditions applied to the harvest by the state forestry authority. Recent advances in and increasing access to electronic means of communication (i.e., e-mail and the Internet) might make these viable survey instruments in the future.

Mail surveys with a telephone follow-up and/or site visit are an efficient method of collecting information. To ensure comparability of results, information collected as part of a self-evaluation—whether collected through the mail, over the phone, or during site visits—should be collected in a manner that does not favor one method over the others. Ideally, telephone follow-up and site visit interviews should consist of no more than reading the questions on the questionnaire, without providing any additional explanation or information that would not have been available to those who responded through the mail. This approach eliminates as much as possible any bias associated with the different means of collecting the information. Questionnaire design is discussed in Section 4.4.3.

It is important that the accuracy of information received through mail and phone surveys be checked. Inaccurate or incomplete responses to questions on mail and/or telephone surveys commonly result from survey respondents misinterpreting questions and thus providing misleading information, not including all relevant information in their responses, not wanting to provide some types of information, or deliberately providing some inaccurate responses. Therefore, the accuracy of information received through mail and phone surveys should be checked by selecting a subsample of the harvesters and/or landowners surveyed and conducting follow-up site visits.

### 4.4.2 Cost

Cost can be an important consideration when selecting an evaluation method. Site visits can cost several hundred dollars per harvest operation depending on the complexity of the operation, the information to be collected, and the number of evaluators used. Mail and/or telephone surveys can be an inexpensive means of collecting information, but their cost must be balanced with the type and accuracy of information that can be collected through them. Other costs need to be figured into the overall cost of mail and/or telephone surveys as well, including follow-up phone calls and site visits to make up for a poor response to mailings and for accuracy checks. Additionally, the cost of questionnaire design must be considered, as a well-designed questionnaire is extremely important to the success of self-evaluations. Questionnaire design is discussed in the next section.

The number of evaluators used for site visits has an obvious impact on the cost of a MM/BMP implementation survey. Survey costs can be minimized by having one or two evaluators visit harvest sites instead of having multiple-person teams visit each survey site. If the expertise of many specialists is desired, it might be cost-effective to have multiple-person teams check the quality of evaluations conducted by one or two evaluators. This can usually be done at a subsample of harvest sites after the sites have been surveyed.

An important factor to consider when determining the number of evaluators to include on site visitation teams, and how to balance the use of one to two evaluators versus multiple-person teams, is the objectives of the survey. Cost notwithstanding, the teams conducting the site evaluations must be sufficient to meet the objectives of the survey, and if the required teams would be too costly, then the objectives of the survey would need to be modified.

Another factor that contributes to the cost of a MM/BMP implementation survey is the number of sites to be surveyed. Once again, a balance must be reached between cost, the objectives of the survey, and the number of sites to be evaluated. Generally, once the objectives of the study have been specified, the number of sites to be evaluated is determined statistically to meet required data quality objectives. If the number of sites that is determined in this way would be too costly, then it would be necessary to modify the study objectives or the data quality objectives. Statistical determination of the number of sites to evaluate is discussed in Section 2.3.

### 4.4.3 Questionnaire Design

Many books have been written on the design of data collection forms and questionnaires (e.g., Churchill, 1983; Ferber et al., 1964; Tull and Hawkins, 1990), and these can provide good advice for the design of simple questionnaires to be used for a single survey. However, for complex questionnaires or ones that will be used for initial surveys as part of a series of surveys (i.e., trend analysis), it is strongly advised that a professional in questionnaire design be consulted. Although it might seem that designing a questionnaire is a simple task, small details such as the order of questions, the selection of one word or phrase over a similar one, and the tone of the questions can significantly affect survey results. A professionally designed questionnaire can yield information beyond that contained in the responses to the questions themselves, while a poorly designed questionnaire can invalidate the results.

The objective of a questionniare, which should be closely related to the objectives of the survey, should be extremely well thought out prior to its being designed. Questionnaires should also be designed at the same time as the information to be collected is selected to ensure that the questions address the objectives as precisely as possible. Conducting these activities simultaneously also provides immediate feedback on the attainability of the objectives and the level of detail of information that can be collected. For example, a researcher might want information on protection of habitat near surface waters, but might discover while designing the questionnaire that the desired information cannot be obtained through the use of a questionnaire, or that the information that could be collected would be insufficient to fully address the chosen objectives. In such a situation the researcher could revise the objectives and questions before going further with questionnaire design.

Tull and Hawkins (1990) identified seven major elements of questionnaire construction:

1. Preliminary decisions
2. Question content
3. Question wording
4. Response format
5. Question sequence
6. Physical characteristics of the questionnaire
7. Pretest and revision

Preliminary decisions include determining exactly what type of information is required, determining the target audience, and selecting the method of communication (e.g, mail, telephone, site visit). These subjects are addressed in other sections of this guidance.

The second step is to determine the content of the questions. Each question should generate one or more of the information requirements identified in the preliminary decisions. The ability of the question to produce the necessary data needs to be assessed. "Double-barreled" questions, in which two or more questions are asked as one, should be avoided. Questions that require the respondent to aggregate several sources of information should be subdivided into several specific questions or parts. The ability of the respondent to answer accurately should also be considered when preparing questions. Some respondents might be unfamiliar with the type of information requested or the terminology used. A respondent might have forgotten some of the information of interest, or might be unable to verbalize an answer. Consideration should be given to the willingness of respondents to answer the questions accurately. If a respondent feels that a particular answer might be embarrassing or personally harmful (e.g., might lead to fines or increased regulation), he or she might refuse to answer the question or might deliberately provide inaccurate information. For this reason, answers to questions that might lead to such responses should be checked for accuracy whenever possible.

The next step is to decide on the specific phrasing of the questions. Simple, easily understood language is preferred. The wording should not bias the answer or be too subjective. For instance, a question should not ask whether groundskidding led to erosion during the harvest. Instead, a series of questions could ask whether groundskidding was used, the slope of land on which it was used, which BMPs were used initially to control erosion from groundskidding, and what additional measures were used to control erosion from groundskidding (if erosion occurred). These questions all request factual information of which a forest operator should be knowledgeable, and the questions progress from simple to more complex. All alternatives and assumptions should be clearly stated on the questionnaire, and the respondent's frame of reference should be considered.

Fourth, the type of response format should be selected. Various types of information can best be obtained using open-ended, multiple-choice, or dichotomous questions. An open-ended question allows respondents to answer in any way they feel is appropriate. Multiple-choice questions tend to reduce some types of bias and are easier to tabulate and analyze; however, good multiple-choice questions can be more difficult to formulate. Dichotomous questions allow only two responses, such as "yes-no" or "agree-disagree." Dichotomous questions are suitable for determining points of fact, but they must be very precisely stated and unequivocally solicit only a single piece of information.

The fifth step in questionnaire design is the ordering of the questions. The first questions should be simple to answer, objective, and interesting in order to relax the respondent. The questionnaire should move from topic to topic in a logical manner without confusing the respondent. Early questions that could bias the respondent should be avoided. There is evidence that response quality declines near the end of a long questionnaire (Tull and Hawkins, 1990). Therefore, more important information should be solicited early. Before presenting the questions, the questionnaire should explain how long (on average) it will take to complete and the types of information that will be solicited. The questionnaire should not present the respondent with any surprises.

The layout of the questionnaire should make it easy to use and should minimize recording mistakes. The layout should clearly show the respondent all possible answers. For mail surveys, an attractive appearance is important for securing cooperation.

The final step in the design of a questionnaire is the pretest and possible revision. A questionnaire should always be pretested with members of the target audience. This will preclude expending large amounts of effort and then discovering that the questionnaire produces biased or incomplete information.

## 4.5 AERIAL RECONNAISSANCE AND PHOTOGRAPHY

Aerial reconnaissance and photography can be useful tools for gathering harvest site information quickly and comparatively inexpensively. For the purposes of forestry BMP compliance surveying, aerial reconnaissance can be useful for selecting survey sites and evaluating some aspects of harvest sites. In Florida, survey sites for each county are selected from fixed-wing aircraft flown in a predetermined pattern. This approach reduces bias in selecting survey sites. The selected sites are then visited by foresters for BMP compliance evaluations (Vowell and Gilpin, 1994). Survey sites are selected from fixed-wing aircraft in South Carolina as well (Adams, 1994). In addition, aerial photography has been proven to be helpful for forest regeneration assessment (Hall and Aldred, 1992; Hudson, 1988); forest inventory and analysis (Hackett, 1988); terrain stratification, riparian area delineation, vegetation mapping, stream morphology characterization, inventory site identification, planning, and monitoring in mountainous regions (Born and Van Hooser, 1988; Hetzel, 1988); rangeland monitoring (BLM, 1991); and agricultural conservation practice identification (Pelletier and Griffin, 1988). Factors such as the characteristics of what is being monitored, scale, and camera format determine how useful aerial photography can be for a particular purpose.

Photographic scale and resolution must be taken into consideration when deciding whether to use aerial photography, and a photographic scale that produces good resolution of the items of importance to the monitoring effort must be chosen. Born and Van Hooser (1988), investigating the usefulness of aerial photography for the classification of inventory and monitoring sample points and locating the sample points on the ground, found that a scale of 1:58,000 (i.e., 1 unit on a photograph represents 58,000 units on the ground) was marginal for use in forestry resource inventorying and monitoring. Hetzel (1988) and Mereszczak (1988), using a large-format camera (see below), found that at a scale of 1:30,000 riparian areas were easily distinguishable and could be delineated with 100 percent accuracy, and cover types could be delineated with 83 percent accuracy. Mereszczak (1988) found that aerial photography was especially useful for monitoring riparian areas because changes in their ecological condition in response to management practices are evident over time frames of 10 years or less. Reutebuch and Shea (1988) reported that photographs taken at a scale of 1:12,000 or larger have a typical resolution of less than 1 foot on the ground. Hall and Aldred (1992) were able to clearly delineate and map cutovers and nonforest areas (water bodies, roads, landings, clearings, brush areas) at a photographic scale of 1:10,000, and could detect conifer seedlings 30 cm or taller, if not hidden beneath other trees, at scales of 1:800 to 1:500. The Bureau of Land Management (BLM) uses low-level, large-scale (1:1,000 to 1:1,500) aerial photography to monitor rangeland vegetation (BLM, 1991). BLM reports that scales smaller than 1:1,500 (e.g., 1:10,000, 1:30,000) are too small to monitor the classes of land cover (shrubs, grasses and forbs, bare soil, rock) on rangeland.

Pelletier and Griffin (1988) investigated the use of aerial photography for the identification of agriculture conservation practices. They found that practices that occupy a large area and have an identifiable pattern, such as contour cropping, strip cropping, terraces, and windbreaks, were readily identified even at a small scale (1:80,000) but that smaller, single-unit practices, such as sediment basins and sediment diversions, were difficult to identify at a small scale. They estimated that 29 percent of practices could be identified at a scale of 1:80,000, 45 percent could be identified at 1:30,000, 70 percent could be identified at 1:15,000, and over 90 percent could be identified at a scale of 1:10,000.

Camera format is another factor that must be considered. Large-format cameras are generally preferred over small-format cameras (e.g., 35mm), but are more costly to purchase and operate. The large negative size (9 cm x 9 cm) produced using a large-format camera provides the resolution and detail necessary for accurate photo interpretation. Large-format cameras can be used from higher altitudes than small-format cameras, and the image area covered by a large-format image at a given scale (e.g., 1:1,500) is much larger than the image area captured by a small-format camera at the same scale. Small-format cameras can be used for identifications that involve large-scale features, such as mining areas, the extent of burning, and large animals in censuses, and they have definite applications in forestry as well. Owens (1988) recommends 35mm photography for monitoring small areas ($\leq 3$ mi$^2$) at low altitude. A particularly useful application of 35mm photography is mapping private landowner parcels and change monitoring (Owens, 1988). Owens (1988) used large-format photographs as baseline data and in subsequent years used 35mm photographs to monitor timber harvests with much success. Small-format cameras are limited in the resolution that they provide when photographs are enlarged (Owens, 1988).

BLM recommends the use of a large-format camera because it provides flexibility to increase sample plot size, it permits modest navigational errors during overflight, and the images provide the photo interpreter with more geographical reference points (BLM, 1991). Large-scale photographs have advantages over topographic maps. Specifically, they have much higher resolution, contain many more features and ground characteristics, and—when viewed in stereo—provide an accurate, 3-dimensional model of an area, complete with vegetative cover information and land-use characteristics (Reutebuch and Shea, 1988). Also, large-format photography equipment is standard equipment for most photo contractors, so one could be hired to take the photographs in lieu of purchasing the equipment.

A drawback to the use of aerial photography is that forestry BMPs that do not meet implementation or operational standards but are similar to practices that do are indistinguishable from ones that do in an aerial photograph (Pelletier and Griffin, 1988). Also, practices that are defined by managerial concepts rather than physical criteria, such as preharvest planning or forest chemical management, cannot be detected with aerial photographs.

Regardless of scale, format, or item being monitored, it is useful for photo interpreters to receive 2 or 3 days of training on the basic fundamentals of photo interpretation and important that they be thoroughly familiar with the vegetation and landforms in the areas where the photographs that they will be interpreting were taken (BLM, 1991). A site visit to the field sites in the photographs is recommended to improve correlation between the interpretation and actual site characteristics. Usually, after a few site visits and interpretations of photographs of those sites, photo interpreters will be familiar with the photographic characteristics of the vegetation in the area and site visits can be reserved for verification of items in doubt. A change in type of vegetation or physiography in photographs normally requires new site visits until photo interpreters are familiar with the characteristics of the new vegetation in the photographs.

# CHAPTER 5. PRESENTATION OF EVALUATION RESULTS

## 5.1 INTRODUCTION

The first three chapters of this guidance presented techniques for the collection of information. Data analysis and interpretation are addressed in detail in Chapter 8 of EPA's *Nonpoint Source Monitoring and Evaluation Guide* (USEPA, 1996). This chapter provides ideas for the presentation of results.

The presentation of MM/BMP compliance evaluation results, whether written or oral, is an integral part of a successful compliance survey. The quality of the presentation of results is an indication of the quality of the survey, and if the presentation fails to convey important information from the compliance survey to those who need the information, the compliance survey itself might be considered a failure.

The quality of the presentation of results is dependent on at least four criteria—results must be complete, accurate, clear, and concise (Churchill, 1983). Completeness means that the presentation provides all necessary information to the audience in language that it understands; accuracy is determined by how well a researcher handles the data, phrases findings, and reasons; clarity is the result of clear and logical thinking and a precision of expression; and conciseness is the result of selecting for inclusion only that which is necessary.

Throughout the process of preparing the results of an MM/BMP compliance survey for presentation, it must be kept in mind that the survey was initially undertaken to provide information for management purposes—specifically, to help make a decision (Tull and Hawkins, 1990). The presentation of results should be built around the decision that the compliance survey was undertaken to support. The message of the presentation must also be tailored to that decision. It must be realized that there will be a time lag between the compliance survey and presentation of the results, and the results should be presented in light of their applicability to the management decision to be made based on them. The length of the time lag is a key factor in determining this applicability. If the time lag is significant, it should be made clear during the presentation that the situation might have changed since the survey was conducted. If reliable trend data are available, the person making the presentation might be able to provide a sense of the likely magnitude of any change in the situation. If the change in status is thought to be insignificant, evidence should be presented to support this claim. For example, state that "At the time that the compliance survey was conducted, forest harvesters were using BMPs with increasing frequency, and the lack of any changes in program implementation coupled with continued interaction with forest harvesters provides no reason to believe that this trend has changed since that time." It would be misleading to state "The monitoring study indicates that forest harvesters *are* using BMPs with increasing frequency." The validity and force of the message will be enhanced further through use of the active voice (*we believe*) rather than the passive voice (*it is believed*).

Three major factors must be considered when presenting the results of MM and BMP implementation studies: (1) identifying the target audience, (2) selecting the appropriate medium (printed word, speech, pictures, etc.), and (3) selecting the most appropriate format to meet the needs of the audience.

## 5.2 AUDIENCE IDENTIFICATION

Identification of the audience(s) to which the results of the MM and BMP implementation study will be presented determines the content and format of the presentation. For the results of implementation monitoring studies, there are typically six potential audiences:

- Interested/concerned citizens
- Forest land owners and harvesters
- Media/general public
- Policy makers
- Resource managers
- Scientists

These audiences have different information needs, interests, and abilities to understand complex data. It is the job of the person(s) preparing the presentation to analyze these factors prior to preparing a presentation. The four criteria for presentation quality apply regardless of the audience. Other elements of a comprehensive presentation, such as discussion of the objectives and limitations of the study and necessary details of the method, must be part of the presentation and must be tailored to the audience. For instance, details of the sampling plan, why the plan was chosen over others, and the statistical methods used for analysis should be recorded even if they are not part of any presentation of results because of their value for future reference when the monitoring is repeated or similar studies are undertaken, but they are best not included in a presentation to management.

## 5.3 PRESENTATION FORMAT

Regardless of whether the results of a monitoring study are presented in writing or orally, or both, the information being presented must be understandable to the audience. Consideration of who the audience is will help ensure that the presentation is particularly suited to the audience's needs. Selection of the correct format for the presentation will ensure that the information is conveyed in a manner that is easy to comprehend.

Most reports will have to be presented in both a written and an oral form. Written reports are valuable for peer review, public information dissemination, and future reference. Oral presentations are often necessary for managers, who usually do not have time to read an entire report, have need for only the results of the study, and are usually not interested in the finer details of the study. Different versions of a report—for the public, scientists, and managers (i.e., an executive summary)—might well have to be written, and separate oral presentations for different audiences—the public, farmers, managers, and scientists at a conference—might have to be prepared.

Most information can most effectively be presented in the form of tables, charts, and diagrams (Tull and Hawkins, 1990). These graphic forms of data and information presentation can help simplify the presentation, making it easier for an audience to comprehend than if explained exhaustively with words. Words are important for pointing out significant ideas or findings, and for interpreting the results where appropriate. Words should not be used to repeat what is already adequately explained in graphics, and slides or transparencies that are composed largely of words should contain only a few essential ideas each. Presentation of too much written information on a single slide or transparency only confuses the audience. Written slides or transparencies should also be free of graphics, such as clever logos or background highlights—unless the pictures are essential to understanding the information presented—since they only make the slides or transparencies more difficult to read. Examples of graphics and written slides are presented in Figures 4-1 through 4-3.

Different types of graphics have different uses as well. Information presented in a tabular format can be difficult to interpret because the reader has to spend some time with the information to extract the essential points from it. The same information presented in a pie chart or bar graph can convey essential information immediately and avoid the inclusion of background data that are not essential to the point. When preparing information for a report, a researcher should organize the information in various ways and choose that which conveys only the information essential for the audience in the least complicated manner.

### 5.3.1 Written Presentations

The following criteria should be considered when preparing written material:

- *Reading level or level of education* of the target audience.
- *Level of detail necessary* to make the results understandable to the target audience. Different audiences require various levels of background information to fully understand the results of a study.
- *Layout*. The integration of text, graphics, color, white space, columns, sidebars, and other design elements is important in the production of material that the target audience will find readable and visually appealing.
- *Graphics*. Photos, drawings, charts, tables, maps, and other graphic elements can be used to effectively present information that the reader might otherwise not understand.

### 5.3.2. Oral Presentations

An effective oral presentation requires special preparation. Tull and Hawkins (1990) recommend three steps:

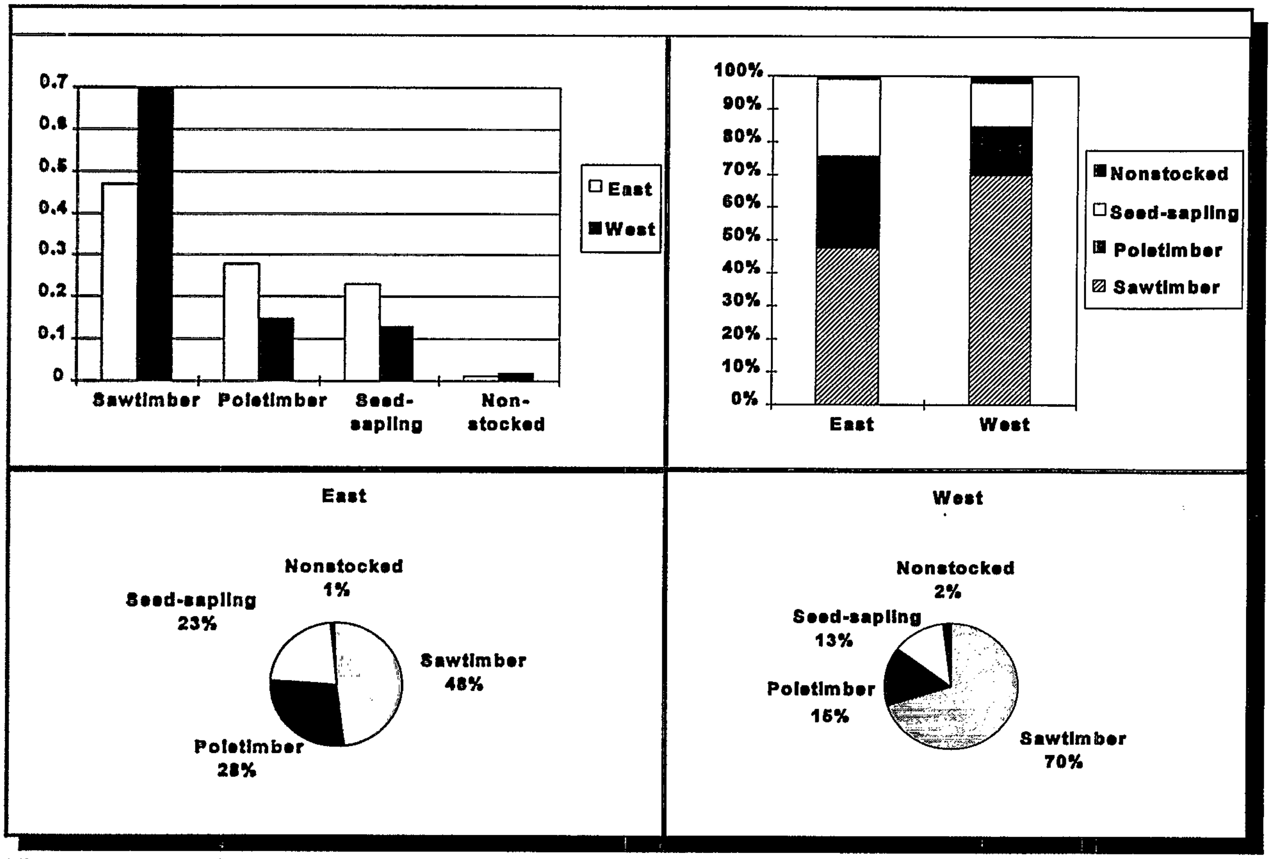



Figure 5-1. Timberland area by stand size class, East and West, 1992, represented graphically in three ways. (After Powell et al., 1994)

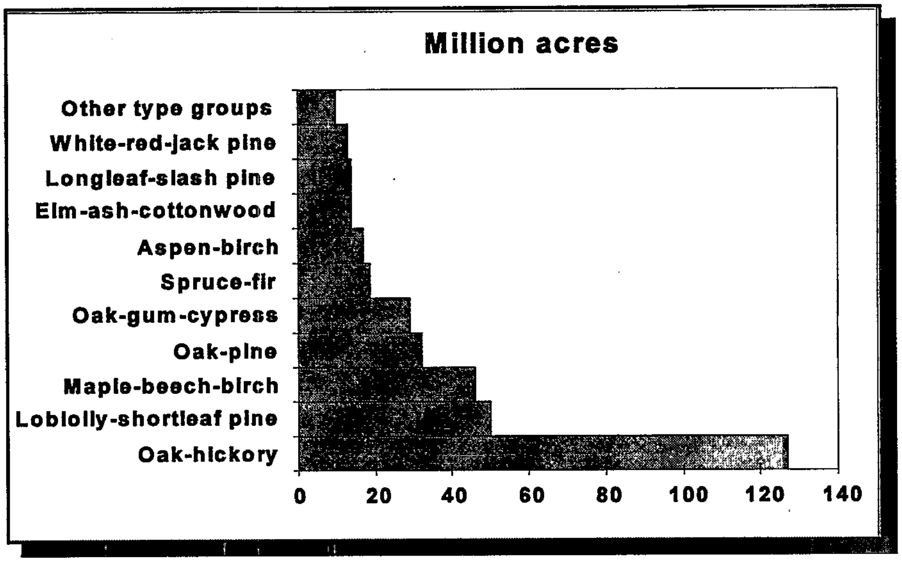



Figure 5-2. Forest type groups on unreserved forest land in the East, 1992, represented graphically. (After Powell et al., 1994)

# FOREST LAND AREA

- Increased 0.1% between 1987 and 1992
- 33% of U.S. land area (737 million acres)
- Clearing forests for agriculture largely halted by 1920
- 34% is federally owned
- 6% is reserved from commercial harvest (47 million acres)

Figure 5-3. Example written presentation slide.

1. Analyze the audience, as explained above.

2. Prepare an outline of the presentation, and preferably a written script.

3. Rehearse the presentation. Several dry runs should be made, and if possible the presentation should be taped on a VCR and analyzed.

These steps are extremely important if an oral presentation is to be effective. Remember that oral presentations of ½ to 1 hour are often all that is available for the presentation of the results of months of research to managers who are poised to make decisions based on the presentation. Adequate preparation is essential if the oral presentation is to accomplish its purpose.

### 5.4 For Further Information

Providing specific examples of effective and ineffective presentation graphics, writing styles, and methods of organization is beyond the scope of this document. A number of resources that contain suggestions for how study results should be presented are available, however, and should be consulted. A listing of some references is provided below.

- *The New York Public Library Writer's Guide to Style and Usage* (NYPL, 1994) has information on design, layout, and presentation in addition to guidance on grammar and style.

- *Good Style: Writing for Science and Technology* (Kirkman, 1992) provides techniques for presenting technical material in a coherent, readable style.

- *The Modern Researcher* (Barzun and Graff, 1992) explains how to turn research into readable, well-organized writing.

- *Writing with Precision: How to Write So That You Cannot Possibly Be Misunderstood*, 6th ed. (Bates, 1993) addresses communication problems of the 1990s.

- *Designer's Guide to Creating Charts & Diagrams* (Holmes, 1991) gives tips for combining graphics with statistical information.

- *The Elements of Graph Design* (Kosslyn, 1993) shows how to create effective displays of quantitative data.

## REFERENCES


Academic Press. 1992. *Dictionary of science and technology*. Academic Press, Inc., San Diego, California.

Adams, T. 1994. *Implementation monitoring of forestry best management practices on harvested sites in South Carolina*. Best Management Practices Monitoring Report BMP-2. South Carolina Forestry Commission, Columbia, South Carolina. September.

Barzun, J., and H.F. Graff. 1992. *The modern researcher*. 5th ed. Houghton Mifflin.

Bates, J. 1993. *Writing with precision: How to write so that you cannot possibly be misunderstood.* 6th ed. Acropolis.

Blalock, H.M., Jr. 1979. *Social statistics*. Rev. 2nd ed. McGraw-Hill Book Company, New York, NY.

BLM. 1991. *Inventory and monitoring coordination: Guidelines for the use of aerial photography in monitoring*. Technical Report TR 1734-1. Department of the Interior, Bureau of Land Management.

Born, J.D., and D.D. Van Hooser. 1988. Intermountain Research Station remote sensing use for resource inventory, planning, and monitoring. In *Remote sensing for resource inventory, planning, and monitoring*. Proceedings of the Second Forest Service Remote Sensing Applications Conference, Sidell, Louisiana, and NSTL, Mississippi, April 11-15, 1988.

Casley, D.J., and D.A. Lury. 1982. *Monitoring and evaluation of agriculture and rural development projects*. The Johns Hopkins University Press, Baltimore, MD.

Churchill, G.A., Jr. 1983. *Marketing research: Methodological foundations,* 3rd ed. The Dryden Press, New York, New York.

Cochran, W.G. 1977. *Sampling techniques*. 3rd ed. John Wiley and Sons, New York, New York.

Colla, J. 1994. *1993 forest practices evaluation report*. Report No. 2-94 FPA. State of Idaho, Department of Lands.

Conover, W.J. 1980. *Practical nonparametric statistics*, 2nd ed. Wiley, New York.

Cross-Smiecinski, A., and L.D. Stetzenback. 1994. *Quality planning for the life science researcher: Meeting quality assurance requirements*. CRC Press, Boca Raton, Florida.

Curtis, J.G., D.W. Pelren, D.B. George, V.D. Adams, and J.B. Layzer. 1990. *Effectiveness of best management practices in preventing degradation of streams caused by silvicultural activities in Pickett State Forest, Tennessee*. Tennessee Department of Conservation, Division of Forestry, Nashville, Tennessee, and Tennessee Wildlife Resources Agency, Nashville, Tennessee. December.

Dissmeyer, G.E. 1994. *Evaluating the effectiveness of forestry best management practices in meeting water quality goals or standards*. Miscellaneous Publication 1520. U.S. Department of Agriculture, Forest Service. July.

Ehinger, W., and D. Potts. 1991. *On-site assessment of best management practices as an indicator of cumulative watershed effects in the Flathead Basin*. Flathead Basin Water Quality and Fisheries Cooperative, University of Montana, School of Forestry, Missoula, Montana.

Ferber, R., D.F. Blankertz, and S. Hollander. 1964. *Marketing research*. The Ronald Press Company, New York, NY.

Freund, J.E. 1973. *Modern elementary statistics*. Prentice-Hall, Englewood Cliffs, New Jersey.

Gaugush, R.F. 1987. *Sampling design for reservoir water quality investigations*. Instruction Report E-87-1. Department of the Army, US Army Corps of Engineers, Washington, DC.

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

Hackett, R.L. 1988. Remote sensing at the North Central Forest Experiment Station. In *Remote sensing for resource inventory, planning, and monitoring*. Proceedings of the Second Forest Service Remote Sensing Applications Conference, Sidell, Louisiana, and NSTL, Mississippi, April 11-15, 1988.

Hall, R.J., and A.H. Alred. 1992. Forest regeneration appraisal with large-scale aerial photographs. *The Forestry Chronicle* 68(1):142-150.

Helsel, D.R., and R.M. Hirsch. 1995. *Statistical methods in water resources*. Elsevier. Amsterdam.

Hetzel, G.E. 1988. Remote sensing applications and monitoring in the Rocky Mountain region. In *Remote sensing for resource inventory, planning, and monitoring*. Proceedings of the Second Forest Service Remote Sensing Applications Conference, Sidell, Louisiana, and NSTL, Mississippi, April 11-15, 1988.

Holmes, N. 1991. *Designer's guide to creating charts & diagrams.* Watson-Guptill.

Hook, D., W. McKee, T. Williams, B. Baker, L. Lundquist, R. Martin, and J. Mills. 1991. *A survey of voluntary compliance of forestry BMPs.* South Carolina Forestry Commission, Columbia, South Carolina.

Hudson, W.D. 1988. Monitoring the long-term effects of silvicultural activities with aerial photography. *J. Forestry* (March):21-26.

IDDHW. 1993. *Forest practices water quality audit 1992.* Idaho Department of Health and Welfare, Division of Environmental Quality, Boise, Idaho.

Kirkman, J. 1992. *Good style: Writing for science and technology.* Chapman and Hall.

Kosslyn, S.M. 1993. *The elements of graph design.* W.H. Freeman.

Kupper, L.L., and K.B. Hafner. 1989. How appropriate are popular sample size formulas? *Am. Stat.* 43:101-105.

MacDonald, L.H., A.W. Smart, and R.C. Wissmar. 1991. *Monitoring guidelines to evaluate the 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.

Mann, H.B., and D.R. Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18:50-60.

McNew, R.W. 1990. Sampling and Estimating Compliance with BMPs. In *Workshop on implementation monitoring of forestry best management practices*, Southern Group of State Foresters, USDA Forest Service, Southern Region, Atlanta, GA, January 23-25, 1990, pp. 86-105.

Mereszczak, I. 1988. Applications of large format camera - color infrared photography to monitoring vegetation management withing the scope of forest plans. In *Remote sensing for resource inventory, planning, and monitoring*. Proceedings of the Second Forest Service Remote Sensing Applications Conference, Sidell, Louisiana, and NSTL, Mississippi, April 11-15, 1988.

NYPL. 1994. *The New York public library writer's guide to style and usage.* A Stonesong Press book. HarperCollins Publishers, New York, NY.

Owens, T. 1988. Using 35mm photographs in resource inventories. In *Remote sensing for resource inventory, planning, and monitoring.* Proceedings of the Second Forest Service Remote Sensing Applications Conference, Sidell, Louisiana, and NSTL, Mississippi, April 11-15, 1988.

Pelletier, R.E., and R.H. Griffin. 1988. An evaluation of photographic scale in aerial photography for identification of conservation practices. *J. Soil Water Conserv.* 43(4):333-337.

Powell, D.S., J.L. Faulkner, D.R. Darr, Z. Zhu, and D.W. MacCleery. 1994. *Forest resources of the United States, 1992.* General Technical Report RM-234 (Revised). U.S. Department of Agriculture, Forest Service. Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado. June.

Rashin, E., C. Clishe, and A. Loch. 1994. *Effectiveness of forest road and timber harvest best management practices with respect to sediment-related water quality impacts.* Interim Report No. 2. Washington State Department of Ecology, Environmental Investigations and Laboratory Services program, Wastershed Assessments Section. Ecology Publication No. 94-67. Olympia, Washington.

Remington, R.D., and M.A. Schork. 1970. *Statistics with applications to the biological and health sciences.* Prentice-Hall, Englewood Cliffs, New Jersey.

Reutebuch, S.E., and R.D. Shea. 1988. A method to control large-scale aerial photos when surveyed ground control is unavailable. In *Remote sensing for resource inventory, planning, and monitoring.* Proceedings of the Second Forest Service Remote Sensing Applications Conference, Sidell, Louisiana, and NSTL, Mississippi, April 11-15, 1988.

Rossman, R., and M.J. Phillips. 1991. *Minnesota forestry best management practices implementation monitoring. 1991 forestry field audit.* Minnesota Department of Natural Resources, Division of Forestry.

Schultz, B. 1992. *Montana forestry best management practices implementation monitoring.* The 1992 Forestry BMP Audits Final Report. Montana Department of State Lands, Forestry Division, Missoula, MT.

Snedecor, G.W. and W.G. Cochran. 1980. *Statistical methods.* 7th ed. The Iowa State University Press, Ames, Iowa.

Tull, D.S., and D.I. Hawkins. 1990. *Marketing research. Measurement and method.* Fifth edition. Macmillan Publishing Company, New York, New York.

USDA. 1992. *Investigating water quality in the pacific southwest region. BMPEP. Best management practices evaluation program: A user's guide.* U.S. Department of Agriculture, Forest Service, Region 5, San Francisco, California. May.

USEPA. 1993a. *Guidance specifying management measures for sources of nonpoint pollution in coastal waters.* EPA 840-B-92-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA. 1993b. *Evaluation of the experimental rural clean water program.* EPA 841-R-93-005. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA. 1995. *National water quality inventory: 1994 report to Congress.* EPA841-R-95-005. U.S. Environmental Protection Agency, Office of Water, Washington, DC. December.

USEPA. 1996. *Nonpoint source monitoring and evaluation guide.* Final review draft. U.S. Environmental Protection Agency, Office of Water, Washington, DC. November.

Vowell, J., and T. Gilpin. 1994. *Florida silviculture best management practices compliance survey report 1993.* Florida Department of Agriculture and Consumer Services, Division of Forestry, Tallahassee, Florida. May.

Wilcoxon, F. 1945. Individual comparisons by ranking metheds. *Biometrics* 1:80-83.

Winer, B.J. 1971. *Statistical principles in experimental design.* McGraw-Hill Book Company, New York.

*accuracy*: the extent to which a measurement approaches the true value of the measured quantity

*aerial photography*: the practice of taking photographs from an airplane, helicopter, or other aviation device while it is airborne

*allocation, Neyman*: stratified sampling in which the cost of sampling each stratum is in proportion to the size of the stratum but variability between strata changes

*allocation, proportional*: stratified sampling in which the variability and cost of sampling for each stratum are in proportion to the size of the stratum

*allowable error*: the level of error acceptable for the purposes of a study

*ANOVA*: see test, analysis of variance

*assumptions*: characteristics of a population of a sampling method taken to be true without proof

*bar graph*: a representation of data wherein data is grouped and represented as vertical or horizontal bars over an axis

*best professional judgement*: an informed opinion made by a professional in the appropriate field of study or expertise

*best management practice*: a practice or combination of practices that are determined to be the most effective and practicable means of controlling point and nonpoint pollutants at levels compatible with environmental quality goals

*bias*: a characteristic of samples such that when taken from a population with a known parameter, their average does not give the parametric value

*binomial*: an algebraic expression that is the sum or difference of two terms

*camera format*: refers to the size of the negative taken by a camera. 35mm is a small camera format

*chi-square distribution*: a scaled quantity whose distribution provides the distribution of the sample variance *coefficient of variation*: a statistical measure used to compare the relative amounts of variation in populations having different means; the standard deviation divided by the mean

*confidence interval*: a range of values about a measured value in which the true value is presumed to lie

*consistency*: conforming to a regular method or style; an approach that keeps all factors of measurement similar from one measurement to the next to the extent possible

*cumulative effects*: the total influences attributable to numerous individual influences

*degrees of freedom*: the number of residuals (the difference between a measured value and the sample average) required to completely determine the others

*design, balanced*: an ANOVA where all cells have equal numbers of samples

*distribution*: the allocation or spread of values of a given parameter among its possible values

*e-mail*: an electronic system for correspondence

*erosion potential*: a measure of the ease with which soil can be carried away in storm water runoff or irrigation runoff

*error*: the fluctuation that occurs from one repetition to another; also *experimental error*

*estimate, baseline*: an appraisal of initial, or actual conditions

*estimate, pooled*: a single estimate obtained from grouping individual estimates and using the latter to obtain a single value

*finite population correction term*: a correction term used when population size is small relative to sample size

*hydrologic modification*: the alteration of the natural circulation or distribution of water by the placement of structures or other activities

*hypothesis, alternative*: the hypothesis which is contrary to the null hypothesis

*hypothesis, null*: the hypothesis or conclusion assumed to be true prior to any analysis

*Internet*: an electronic data transmission system *management measure*: an economically achievable measure for the control of the addition of pollutants from existing and new categories and classes of nonpoint sources of pollution, which reflect the greatest degree of pollutant reduction achievable through the application of the best available nonpoint pollution control practices, technologies, processes, siting criteria, operating methods, or other alternatives

*mean, estimated*: a value of population mean arrived at through sampling

*mean, overall*: the measured average of a population

*mean, stratum*: the measured average within a sample subgroup or stratum

*measurement bias*: a consistent under- or overestimation of the true value of something being measured, often due to the method of measurement

*measurement error*: the deviation of a measurement from the true value of that which is being measured

*median*: the value of the middle term when data are arranged in order of size; a measure of central tendency

*monitoring, baseline*: monitoring conducted to establish initial knowledge about the actual state of a population

*monitoring, compliance*: monitoring conducted to determine if those who must implement programs, best management practices, or management measures, or who must conduct operations according to standards or specifications are doing so

*monitoring, project*: monitoring conducted to determine the impact of a project, activity, or program

*monitoring, validation*: monitoring conducted to determine how well a model accurately reflects reality

*navigational error*: errors in determining the actual location (altitude or latitude/longitude) of an airplane or other aviation device due to instrumentation or the operator

*nominal*: referred to by name; variables that cannot be measured but must be expressed qualitatively *nonparametric method*: distribution-free method; any of various inferential procedures whose conclusions do not rely on assumptions about the distribution of the population of interest

*normal approximation*: an assumption that a population has the characteristics of a normally-distributed population

*normal deviate*: deviation from the mean expressed in units of σ

*nutrient management plan*: a plan for managing the quantity of nutrients applied to crops to achieve maximum plant nutrition and minimum nutrient waste

*ordinal*: ordered such that the position of an element in a series is specified

*parametric method*: any statistical method whose conclusions rely on assumptions about the distribution of the population of interest

*physiography*: a description of the surface features of the Earth; a description of land forms

*pie chart*: a representation of data wherein data is grouped and represented as more or less triangular sections of a circle and the total is the entire circle

*population, sample*: the members of a population that are actually sampled or measured

*population, target*: the population about which inferences are made; the group of interest, from which samples are taken

*population unit*: an individual member of a target population that can be measured independently of other members

*power*: the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true

*precision*: a measure of the similarity of individual measurements of the same population

*question, dichotomous*: a question that allows for only two responses, such as "yes" and "no"

*question, double-barreled*: two questions asked as a single question

*question, multiple-choice*: a question with two or more predetermined responses

*question, open-ended*: a question format that requires a response beyond "yes" or "no" *remote sensing*: methods of obtaining data from a location distant from the object being measured, such as from an airplane or satellite

*resolution*: the sharpness of a photograph

*sample size*: the number of population units measured

*sampling, cluster*: sampling in which small groups of population units are selected for sampling and each unit in each selected group is measured

*sampling, simple random*: sampling in which each unit of the target population has an equal chance of being selected

*sampling, stratified random*: sampling in which the target population is divided into separate subgroups, each of which is more internally similar than the overall population is, prior to sample selection

*sampling, systematic*: sampling in which population units are chosen in accordance with a predetermined sample selection system

*sampling error*: error attributable to actual variability in population units not accounted for by the sampling method

*scale (aerial photography)*: the proportion of the image size of an object (such as a land area) to its actual size, e.g., 1:3,000; the smaller the second number, the larger the scale

*scale system*: a system for ranking measurements or members of a population on a scale, such as 1 to 5

*significance level*: in hypothesis testing, the probability of rejecting a hypothesis that is correct, that is, the probability of a Type I error

*standard deviation*: a measure of spread; the positive square root of the variance

*standard error*: an estimate of the standard deviation of means that would be expected if a collection of means based on equal-sized samples of *n* items from the same population were obtained

*statistical inference*: conclusions drawn about a population using statistics

*statistics, descriptive*: measurements of population characteristics designed to summarize important features of a data set

*stratification*: the process of dividing a population into internally similar subgroups

*stratum*: one of the subgroups created prior to sampling in stratified random sampling

*streamside management area*: a designated area that consists of a water body (e.g., stream) and an adjacent area of varying width where management practices that might affect water quality, fish, or other aquatic resources are modified to protect the water body and its adjacent resources and to reduce the pollution effect of an activity on the water body

*subjectivity*: a characteristic of analysis that requires personal judgement on the part of the person doing the analysis

*target audience*: the population that a monitoring effort is intended to measure

*test, analysis of variance*: a statistical test used to determine whether two or more sample means could have been obtained from populations with the same parametric mean

*test, Friedman*: a nonparametric test that can be used for analysis when two variables are involved

*test, Kruskal-Wallis*: a nonparametric test recommended for the general case with *a* samples and $n_i$ variates per sample

*test, Mann-Whitney*: a nonparametric test for use when a test is only between two samples

*test, student's t*: a statistical test used to test for significant differences between means when only two samples are involved

*test, Tukey's*: a test to ascertain whether the interaction found in a given set of data can be explained in terms of multiplicative main effects

*test, Wilcoxon's*: a nonparametric test for use when a test is only between two samples

*total maximum daily load*: a total allowable addition of pollutants from all affecting sources to an individual water body over a 24-hour period

*transformation, data*: manipulation of data such that it will meet the assumptions required for analysis *unit sampling cost*: the cost of attributable to sampling a single population unit

*variance*: a measure of the spread of data around the mean

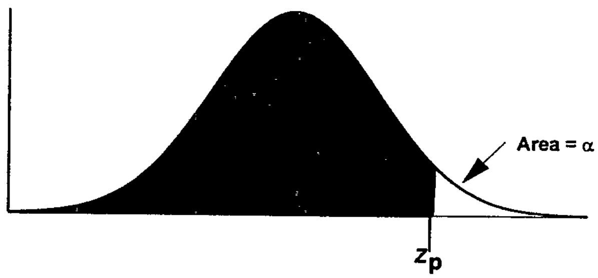
*waterbar*: a diversion ditch and/or hump installed across a trail or road to divert runoff from the surface before the flow gains enough volume and velocity to cause soil movement and erosion

*watershed assessment*: an investigation of numerous characteristics of a watershed in order to describe its actual condition

# *APPENDIX A*

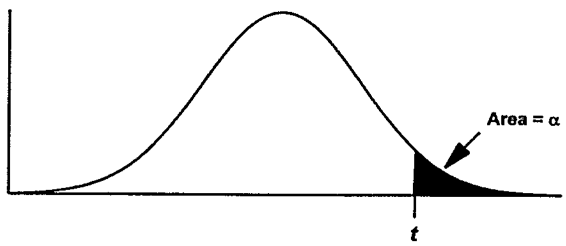
## *Statistical Tables*

**Table A1. Cumulative areas under the Normal distribution** (values of $p$ corresponding to $Z_p$)



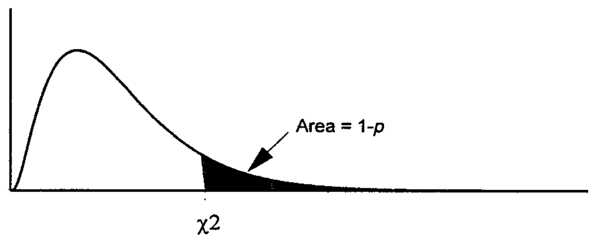

| $Z_p$ | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |
| 2.0 | 0.9772 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 | 0.9817 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 | 0.9857 |
| 2.2 | 0.9861 | 0.9864 | 0.9868 | 0.9871 | 0.9875 | 0.9878 | 0.9881 | 0.9884 | 0.9887 | 0.9890 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 | 0.9916 |
| 2.4 | 0.9918 | 0.9920 | 0.9922 | 0.9925 | 0.9927 | 0.9929 | 0.9931 | 0.9932 | 0.9934 | 0.9936 |
| 2.5 | 0.9938 | 0.9940 | 0.9941 | 0.9943 | 0.9945 | 0.9946 | 0.9948 | 0.9949 | 0.9951 | 0.9952 |
| 2.6 | 0.9953 | 0.9955 | 0.9956 | 0.9957 | 0.9959 | 0.9960 | 0.9961 | 0.9962 | 0.9963 | 0.9964 |
| 2.7 | 0.9965 | 0.9966 | 0.9967 | 0.9968 | 0.9969 | 0.9970 | 0.9971 | 0.9972 | 0.9973 | 0.9974 |
| 2.8 | 0.9974 | 0.9975 | 0.9976 | 0.9977 | 0.9977 | 0.9978 | 0.9979 | 0.9979 | 0.9980 | 0.9981 |
| 2.9 | 0.9981 | 0.9982 | 0.9982 | 0.9983 | 0.9984 | 0.9984 | 0.9985 | 0.9985 | 0.9986 | 0.9986 |
| 3.0 | 0.9987 | 0.9987 | 0.9987 | 0.9988 | 0.9988 | 0.9989 | 0.9989 | 0.9989 | 0.9990 | 0.9990 |
| 3.1 | 0.9990 | 0.9991 | 0.9991 | 0.9991 | 0.9992 | 0.9992 | 0.9992 | 0.9992 | 0.9993 | 0.9993 |
| 3.2 | 0.9993 | 0.9993 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9995 | 0.9995 | 0.9995 |
| 3.3 | 0.9995 | 0.9995 | 0.9995 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9997 |
| 3.4 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9998 |

**Table A2. Percentiles of the $t_{\alpha,df}$ distribution** (values of $t$ such that $100(1-\alpha)\%$ of the distribution is less than $t$)




| df | α = 0.40 | α = 0.30 | α = 0.20 | α = 0.10 | α = 0.05 | α = 0.025 | α = 0.010 | α = 0.005 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.3249 | 0.7265 | 1.3764 | 3.0777 | 6.3137 | 12.7062 | 31.8210 | 63.6559 |
| 2 | 0.2887 | 0.6172 | 1.0607 | 1.8856 | 2.9200 | 4.3027 | 6.9645 | 9.9250 |
| 3 | 0.2767 | 0.5844 | 0.9785 | 1.6377 | 2.3534 | 3.1824 | 4.5407 | 5.8408 |
| 4 | 0.2707 | 0.5686 | 0.9410 | 1.5332 | 2.1318 | 2.7765 | 3.7469 | 4.6041 |
| 5 | 0.2672 | 0.5594 | 0.9195 | 1.4759 | 2.0150 | 2.5706 | 3.3649 | 4.0321 |
| 6 | 0.2648 | 0.5534 | 0.9057 | 1.4398 | 1.9432 | 2.4469 | 3.1427 | 3.7074 |
| 7 | 0.2632 | 0.5491 | 0.8960 | 1.4149 | 1.8946 | 2.3646 | 2.9979 | 3.4995 |
| 8 | 0.2619 | 0.5459 | 0.8889 | 1.3968 | 1.8595 | 2.3060 | 2.8965 | 3.3554 |
| 9 | 0.2610 | 0.5435 | 0.8834 | 1.3830 | 1.8331 | 2.2622 | 2.8214 | 3.2498 |
| 10 | 0.2602 | 0.5415 | 0.8791 | 1.3722 | 1.8125 | 2.2281 | 2.7638 | 3.1693 |
| 11 | 0.2596 | 0.5399 | 0.8755 | 1.3634 | 1.7959 | 2.2010 | 2.7181 | 3.1058 |
| 12 | 0.2590 | 0.5386 | 0.8726 | 1.3562 | 1.7823 | 2.1788 | 2.6810 | 3.0545 |
| 13 | 0.2586 | 0.5375 | 0.8702 | 1.3502 | 1.7709 | 2.1604 | 2.6503 | 3.0123 |
| 14 | 0.2582 | 0.5366 | 0.8681 | 1.3450 | 1.7613 | 2.1448 | 2.6245 | 2.9768 |
| 15 | 0.2579 | 0.5357 | 0.8662 | 1.3406 | 1.7531 | 2.1315 | 2.6025 | 2.9467 |
| 16 | 0.2576 | 0.5350 | 0.8647 | 1.3368 | 1.7459 | 2.1199 | 2.5835 | 2.9208 |
| 17 | 0.2573 | 0.5344 | 0.8633 | 1.3334 | 1.7396 | 2.1098 | 2.5669 | 2.8982 |
| 18 | 0.2571 | 0.5338 | 0.8620 | 1.3304 | 1.7341 | 2.1009 | 2.5524 | 2.8784 |
| 19 | 0.2569 | 0.5333 | 0.8610 | 1.3277 | 1.7291 | 2.0930 | 2.5395 | 2.8609 |
| 20 | 0.2567 | 0.5329 | 0.8600 | 1.3253 | 1.7247 | 2.0860 | 2.5280 | 2.8453 |
| 21 | 0.2566 | 0.5325 | 0.8591 | 1.3232 | 1.7207 | 2.0796 | 2.5176 | 2.8314 |
| 22 | 0.2564 | 0.5321 | 0.8583 | 1.3212 | 1.7171 | 2.0739 | 2.5083 | 2.8188 |
| 23 | 0.2563 | 0.5317 | 0.8575 | 1.3195 | 1.7139 | 2.0687 | 2.4999 | 2.8073 |
| 24 | 0.2562 | 0.5314 | 0.8569 | 1.3178 | 1.7109 | 2.0639 | 2.4922 | 2.7970 |
| 25 | 0.2561 | 0.5312 | 0.8562 | 1.3163 | 1.7081 | 2.0595 | 2.4851 | 2.7874 |
| 26 | 0.2560 | 0.5309 | 0.8557 | 1.3150 | 1.7056 | 2.0555 | 2.4786 | 2.7787 |
| 27 | 0.2559 | 0.5306 | 0.8551 | 1.3137 | 1.7033 | 2.0518 | 2.4727 | 2.7707 |
| 28 | 0.2558 | 0.5304 | 0.8546 | 1.3125 | 1.7011 | 2.0484 | 2.4671 | 2.7633 |
| 29 | 0.2557 | 0.5302 | 0.8542 | 1.3114 | 1.6991 | 2.0452 | 2.4620 | 2.7564 |
| 30 | 0.2556 | 0.5300 | 0.8538 | 1.3104 | 1.6973 | 2.0423 | 2.4573 | 2.7500 |
| 35 | 0.2553 | 0.5292 | 0.8520 | 1.3062 | 1.6896 | 2.0301 | 2.4377 | 2.7238 |
| 40 | 0.2550 | 0.5286 | 0.8507 | 1.3031 | 1.6839 | 2.0211 | 2.4233 | 2.7045 |
| 50 | 0.2547 | 0.5278 | 0.8489 | 1.2987 | 1.6759 | 2.0086 | 2.4033 | 2.6778 |
| 60 | 0.2545 | 0.5272 | 0.8477 | 1.2958 | 1.6706 | 2.0003 | 2.3901 | 2.6603 |
| 80 | 0.2542 | 0.5265 | 0.8461 | 1.2922 | 1.6641 | 1.9901 | 2.3739 | 2.6387 |
| 100 | 0.2540 | 0.5261 | 0.8452 | 1.2901 | 1.6602 | 1.9840 | 2.3642 | 2.6259 |
| 150 | 0.2538 | 0.5255 | 0.8440 | 1.2872 | 1.6551 | 1.9759 | 2.3515 | 2.6090 |
| 200 | 0.2537 | 0.5252 | 0.8434 | 1.2858 | 1.6525 | 1.9719 | 2.3451 | 2.6006 |
| Inf. | 0.2533 | 0.5244 | 0.8416 | 1.2816 | 1.6449 | 1.9600 | 2.3264 | 2.5758 |

**Table A3. Upper and lower percentiles of the Chi-square distribution**




| *df* | *p* = 0.001 | 0.005 | 0.010 | 0.025 | 0.050 | 0.100 | 0.900 | 0.950 | 0.975 | 0.990 | 0.995 | 0.999 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | .... | .... | .... | 0.001 | 0.004 | 0.016 | 2.706 | 3.841 | 5.024 | 6.635 | 7.879 | 10.827 |
| 2 | 0.002 | 0.010 | 0.020 | 0.051 | 0.103 | 0.211 | 4.605 | 5.991 | 7.378 | 9.210 | 10.597 | 13.815 |
| 3 | 0.024 | 0.072 | 0.115 | 0.216 | 0.352 | 0.584 | 6.251 | 7.815 | 9.348 | 11.345 | 12.838 | 16.266 |
| 4 | 0.091 | 0.207 | 0.297 | 0.484 | 0.711 | 1.064 | 7.779 | 9.488 | 11.143 | 13.277 | 14.860 | 18.466 |
| 5 | 0.210 | 0.412 | 0.554 | 0.831 | 1.145 | 1.610 | 9.236 | 11.070 | 12.832 | 15.086 | 16.750 | 20.515 |
| 6 | 0.381 | 0.676 | 0.872 | 1.237 | 1.635 | 2.204 | 10.645 | 12.592 | 14.449 | 16.812 | 18.548 | 22.457 |
| 7 | 0.599 | 0.989 | 1.239 | 1.690 | 2.167 | 2.833 | 12.017 | 14.067 | 16.013 | 18.475 | 20.278 | 24.321 |
| 8 | 0.857 | 1.344 | 1.647 | 2.180 | 2.733 | 3.490 | 13.362 | 15.507 | 17.535 | 20.090 | 21.955 | 26.124 |
| 9 | 1.152 | 1.735 | 2.088 | 2.700 | 3.325 | 4.168 | 14.684 | 16.919 | 19.023 | 21.666 | 23.589 | 27.877 |
| 10 | 1.479 | 2.156 | 2.558 | 3.247 | 3.940 | 4.865 | 15.987 | 18.307 | 20.483 | 23.209 | 25.188 | 29.588 |
| 11 | 1.834 | 2.603 | 3.053 | 3.816 | 4.575 | 5.578 | 17.275 | 19.675 | 21.920 | 24.725 | 26.757 | 31.264 |
| 12 | 2.214 | 3.074 | 3.571 | 4.404 | 5.226 | 6.304 | 18.549 | 21.026 | 23.337 | 26.217 | 28.300 | 32.909 |
| 13 | 2.617 | 3.565 | 4.107 | 5.009 | 5.892 | 7.041 | 19.812 | 22.362 | 24.736 | 27.688 | 29.819 | 34.527 |
| 14 | 3.041 | 4.075 | 4.660 | 5.629 | 6.571 | 7.790 | 21.064 | 23.685 | 26.119 | 29.141 | 31.319 | 36.124 |
| 15 | 3.483 | 4.601 | 5.229 | 6.262 | 7.261 | 8.547 | 22.307 | 24.996 | 27.488 | 30.578 | 32.801 | 37.698 |
| 16 | 3.942 | 5.142 | 5.812 | 6.908 | 7.962 | 9.312 | 23.542 | 26.296 | 28.845 | 32.000 | 34.267 | 39.252 |
| 17 | 4.416 | 5.697 | 6.408 | 7.564 | 8.672 | 10.085 | 24.769 | 27.587 | 30.191 | 33.409 | 35.718 | 40.791 |
| 18 | 4.905 | 6.265 | 7.015 | 8.231 | 9.390 | 10.865 | 25.989 | 28.869 | 31.526 | 34.805 | 37.156 | 42.312 |
| 19 | 5.407 | 6.844 | 7.633 | 8.907 | 10.117 | 11.651 | 27.204 | 30.144 | 32.852 | 36.191 | 38.582 | 43.819 |
| 20 | 5.921 | 7.434 | 8.260 | 9.591 | 10.851 | 12.443 | 28.412 | 31.410 | 34.170 | 37.566 | 39.997 | 45.314 |
| 21 | 6.447 | 8.034 | 8.897 | 10.283 | 11.591 | 13.240 | 29.615 | 32.671 | 35.479 | 38.932 | 41.401 | 46.796 |
| 22 | 6.983 | 8.643 | 9.542 | 10.982 | 12.338 | 14.041 | 30.813 | 33.924 | 36.781 | 40.289 | 42.796 | 48.268 |
| 23 | 7.529 | 9.260 | 10.196 | 11.689 | 13.091 | 14.848 | 32.007 | 35.172 | 38.076 | 41.638 | 44.181 | 49.728 |
| 24 | 8.085 | 9.886 | 10.856 | 12.401 | 13.848 | 15.659 | 33.196 | 36.415 | 39.364 | 42.980 | 45.558 | 51.179 |
| 25 | 8.649 | 10.520 | 11.524 | 13.120 | 14.611 | 16.473 | 34.382 | 37.652 | 40.646 | 44.314 | 46.928 | 52.619 |
| 26 | 9.222 | 11.160 | 12.198 | 13.844 | 15.379 | 17.292 | 35.563 | 38.885 | 41.923 | 45.642 | 48.290 | 54.051 |
| 27 | 9.803 | 11.808 | 12.878 | 14.573 | 16.151 | 18.114 | 36.741 | 40.113 | 43.195 | 46.963 | 49.645 | 55.475 |
| 28 | 10.391 | 12.461 | 13.565 | 15.308 | 16.928 | 18.939 | 37.916 | 41.337 | 44.461 | 48.278 | 50.994 | 56.892 |
| 29 | 10.986 | 13.121 | 14.256 | 16.047 | 17.708 | 19.768 | 39.087 | 42.557 | 45.722 | 49.588 | 52.335 | 58.301 |
| 30 | 11.588 | 13.787 | 14.953 | 16.791 | 18.493 | 20.599 | 40.256 | 43.773 | 46.979 | 50.892 | 53.672 | 59.702 |
| 35 | 14.688 | 17.192 | 18.509 | 20.569 | 22.465 | 24.797 | 46.059 | 49.802 | 53.203 | 57.342 | 60.275 | 66.619 |
| 40 | 17.917 | 20.707 | 22.164 | 24.433 | 26.509 | 29.051 | 51.805 | 55.758 | 59.342 | 63.691 | 66.766 | 73.403 |
| 50 | 24.674 | 27.991 | 29.707 | 32.357 | 34.764 | 37.689 | 63.167 | 67.505 | 71.420 | 76.154 | 79.490 | 86.660 |
| 60 | 31.738 | 35.534 | 37.485 | 40.482 | 43.188 | 46.459 | 74.397 | 79.082 | 83.298 | 88.379 | 91.952 | 99.608 |
| 70 | 39.036 | 43.275 | 45.442 | 48.758 | 51.739 | 55.329 | 85.527 | 90.531 | 95.023 | 100.43 | 104.21 | 112.32 |
| 80 | 46.520 | 51.172 | 53.540 | 57.153 | 60.391 | 64.278 | 96.578 | 101.88 | 106.63 | 112.33 | 116.32 | 124.84 |
| 90 | 54.156 | 59.196 | 61.754 | 65.647 | 69.126 | 73.291 | 107.57 | 113.15 | 118.14 | 124.12 | 128.30 | 137.21 |
| 100 | 61.918 | 67.328 | 70.065 | 74.222 | 77.929 | 82.358 | 118.50 | 124.34 | 129.56 | 135.81 | 140.17 | 149.45 |
| 200 | 143.84 | 152.24 | 156.43 | 162.73 | 168.28 | 174.84 | 226.02 | 233.99 | 241.06 | 249.45 | 255.26 | 267.54 |

# APPENDIX B

## Sample Evaluation Forms

# Best Management Practices Evaluation

UTM Coordinates Zone __ __
Easting __ __ __ __ __ __
Northing __ __ __ __ __ __ __

## Form E08: Road Surface, Drainage and Slope Protection
(BMP 2.2, 2.4, 2.5, 2.7, 2.10, 2.23)

ID#: ____________

Selection Code: ____________

Reviewer(s) ____________ Title(s) ____________ Date ________ Forest ________ District ________ T ____ R S

Project ____________ Road # ________ Year Construction Completed ____________ Last Maintenance ________

Project is: ☐ Construction ☐ Reconstruction ☐ Maintenance ☐ Other (describe) ____________ NFS Watershed

**Implementation**

**Rating**

*1 = Exceeds contract/project requirements*
*2 = Meets contract/project requirements*
*3 = Minor departure from contract/project requirements*
*4 = Major departure from contract/project requirements*
*Rate as NA if criteria not applicable at this site*

For construction or reconstruction projects:

1) Design objectives developed that address water quality issues identified by ID or review team ........ ______
2) Design meets objectives ........ ______
3) Construction/Reconstruction contract requirements met for:
   a) Surfacing ........ ______
   b) Drainage ........ ______
   c) Slope stabilization ........ ______
   d) Slash disposal ........ ______

For maintenance projects:

1) Check appropriate means of maintenance accomplishment: ☐ Timber sale contract
   ☐ Force account
   ☐ Maintenance contract
   ☐ Other ( ____________ )
2) Maintenance specifications were met for:
   a) Surface blading/repair/treatment ........ ______
   b) Drainage structure repair/treatment ........ ______
   c) Slope treatment/sidecast ........ ______

If any rating is "3" or "4", complete the following:

Problem occurred in which phase(s) of the project: ☐ Location ☐ Design ☐ EA ☐ Contract ☐ Construction ☐ Maintenance

Describe deficiencies and corrective actions:

Form #E08 Version 4/92

SITE NUMBER:______________ DATE:______________
OWNERSHIP:______________ OPERATOR:______________
LEGAL DESCRIPTION:______________ SALE OR PROJECT NUMBER:______________
PROJECT ACRES REVIEWED:______________ TEAM INITIALS:______________

| SITE CONDITIONS | PRACTICES |
|---|---|
| LANDFORM:______________<br>GENERAL SOILS:______________<br>DRAINAGE:______________<br>SLOPE RANGE:______________<br>WATER BODIES PRESENT (type):______________<br>______________<br>______________<br><br>DEPTH/WIDTH OF STREAMS(type):______________<br>______________<br>______________<br><br>OTHER:______________<br>______________ | STAGE ("x" if completed)<br>PREHARVEST ( ) ROAD CONSTRUCTION ( )<br>HARVEST ( ) SLASH DISPOSAL ( )<br>SITE PREP ( )<br>DATE OF ACTIVITY______________<br>ROADS:<br>NEW CONSTRUCTION (length):______________<br>RECONSTRUCTION (length):______________<br>HARVEST ACRES:______________<br>HARVEST METHOD:______________<br>SITE PREP ACRES:______________<br>SITE PREP METHOD:______________<br>SLASH DISPOSAL:______________<br>PESTICIDES USED:______________<br>OTHER:______________<br>______________ |

## RATING GUIDE

### APPLICATION

5--OPERATION EXCEEDS REQUIREMENT OF BMP
4--OPERATION MEETS REQUIREMENT OF BMP
3--MINOR DEPARTURE FROM BMP
2--MAJOR DEPARTURE FROM BMP
1--GROSS NEGLECT OF BMP

### EFFECTIVENESS

6--IMPROVED PROTECTION OF SOIL AND WATER RESOURCES OVER PRE-PROJECT CONDITION.
5--ADEQUATE PROTECTION OF SOIL AND WATER RESOURCES.
4--MINOR AND TEMPORARY IMPACTS ON SOIL AND WATER RESOURCES.
3--MAJOR AND TEMPORARY IMPACTS ON SOIL AND WATER RESOURCES.
2--MINOR AND PROLONGED IMPACTS ON SOIL AND WATER RESOURCES.
1--MAJOR AND PROLONGED IMPACTS ON SOIL AND WATER RESOURCES.

### DEFINITIONS (BY EXAMPLE)

ADEQUATE: Small amount of material eroded; material does not reach drainages, streams, lakes or wetlands
MINOR: Erosion and delivery of material to drainages but not to streams, lakes or open-water wetlands.
MAJOR: Erosion and subsequent delivery of sediment to streams, lakes or open water wetlands.
TEMPORARY: Impacts lasting one year or less; no more than one runoff season.
PROLONGED: Impacts lasting more than one year.
* It is possible to have a departure from BMPs and still adequate protection.

| RECOMMENDED BEST MANAGEMENT PRACTICES | APPLICABLE TO SITE (Y/N) | APPLICATION | EFFECTIVENESS | COMMENTS |
|---|---|---|---|---|
| **MECHANICAL SITE PREP** | | | | |
| 15 General Recommendations (p50) | | | | |
| 15a Site prep technique appropriate to the site | | | | |
| 15b Provide adequate filter strips | | | | |
| 15c Avoid operating during periods of saturated soil | | | | |
| 15d Maintain adequate vegetation adjacent to designated trout streams | | | | |
| 15e Site prep technique properly employed (p50-52) | | | | |
| - Shearing and raking | | | | |
| - Disking | | | | |
| - Patch or row scarification | | | | |
| - Other | | | | |
| **PESTICIDE USE** | | | | |
| 16 Prevent entry of pesticide residues into surface and ground waters (p57-75) | | | | |
| **PRESCRIBED BURNING** | | | | |
| 17 Planning (p78) | | | | |
| 17a Obtain proper permits | | | | |
| 18 Prescriptions (p79-81) | | | | |
| 18a Locate fire lines on the contour | | | | |
| 18b Use natural or in-place fire barriers | | | | |
| 18c Establish filter strips for fire lines | | | | |
| 18d Avoid placement of debris piles for burning in filter strips or sensitive areas | | | | |
| 18e Limit water quality impacts from fire line construction by using mowing, herbicides, retardant etc. | | | | |
| 19 Maintenance (p81) | | | | |
| 19a Maintain erosion control measures on firelines | | | | |

SITE ID NO:

# TEXAS BMP MONITORING CHECKLIST

## GENERAL

1. County__________ 2. Block/Grid__________
3. Latitude__________ Longitude__________
Forester: 4. ________ 5.__________
6. Timber Buyer __________
7. Logger __________
8. Activity __________
9. Estimated date of activity __________
10. Acres affected ________
11. Inspector __________

LANDOWNER:
12. Owner Type: N L A I P
13. Name __________
14. Address __________
15. City __________ ZIP ________
16. Phone __________
17. Date of Inspection __________
18. Accompanied by:__________

## SITE CHARACTERISTICS

19. Terrain: F H S
20. Erodability hazard: L M H
21. Type stream present P I
22. Distance to nearest permanent water body: <300' 300-800' 800-1600' 1600'+
23. Predominant soil series/texture: ____ / C CL L SL S

## PERMANENT ROADS

[ ] NOT APPLICABLE

| | |
|---|---|
| 24. Avoid sensitive areas. | Y N NA |
| 25. Roads meet grade specs. | Y N NA |
| 26. Stabilized stream crossing. | Y N NA |
| 27. Rutting within allowable specs. | Y N NA |
| 28. Ditches do not dump into streams. | Y N NA |
| 29. Were BMP's used. | Y N NA |
| Type: RD WD WB RE OC PL RS CU BR LW | |
| 30. Were BMP's effective. | Y N NA |
| 31. Stream free of sediment. | Y N NA |

## SKID TRAILS / TEMPORARY ROADS

[ ] NOT APPLICABLE

| | |
|---|---|
| 32. Slopes less than 15%. | Y N NA |
| 33. Rutting within allowable specs. | Y N NA |
| 34. Water bars evident. | Y N NA |
| 35. Water bars working. | Y N NA |
| 36. Stream crossings minimized. | Y N NA |
| 37. Stream crossings correct. | Y N NA |
| 38. Stream crossings restored & stabilized. | Y N NA |
| 39. Were BMP's used. | Y N NA |
| Type: RD WD WB RE OC PL RS CU BR LW | |
| 40. Stream free of sediment. | Y N NA |

## SMZ

[ ] NOT APPLICABLE

| | |
|---|---|
| 41. SMZ present on permanent stream. | Y N NA |
| 42. SMZ present on intermittent stream. | Y N NA |
| 43. SMZ adequately wide. | Y N NA |
| 44. Thinning within allowable specs. | Y N NA |
| 45. SMZ integrity honored. | Y N NA |
| 46. Stream clear of debris. | Y N NA |
| 47. SMZ free of roads and landings. | Y N NA |
| 48. Stream free of sediment. | Y N NA |

## SITE PREPARATION

[ ] NOT APPLICABLE

| | |
|---|---|
| 49. Site prep method ________ | |
| 50. Regeneration method ________ | |
| 51. No soil movement on site. | Y N NA |
| 52. Firebreak erosion controlled. | Y N NA |
| 53. SMZ integrity honored. | Y N NA |
| 54. Windrows on contour / free of soil. | Y N NA |
| 55. No chemicals off site. | Y N NA |
| 56. Were BMP's used. | Y N NA |
| Type: WB RE OC RS | |
| 57. Stream free of sediment. | Y N NA |

## LANDINGS

[ ] NOT APPLICABLE

| | |
|---|---|
| 58. Locations free of oil / trash. | Y N NA |
| 59. Located outside SMZ. | Y N NA |
| 60. Well drained location | Y N |
| 61. Restored, stabilized. | Y N NN |

62. Overall compliance with Best Management Practices

| NEEDS IMPROVEMENT | | PASS | | |
|---|---|---|---|---|
| NO EFFORT | POOR | FAIR | GOOD | EXCELLENT |

See Evaluation Criteria for a full description of numbered questions.