DRAFT

SAN DIEGO STORMWATER COPERMITTEES BMP EFFECTIVENESS ASSESSMENT STRATEGIES MODEL PROGRAM GUIDANCE

Prepared for:
The County of San Diego
Clean Water Program
Department of Environmental Health
Land and Water Quality Division
1255 Imperial Avenue
San Diego, California 92112

Prepared By:
MEC Analytical Systems, Inc.
2433 Impala Drive
Carlsbad, California 92008

November 2001

TABLE OF CONTENTS

1.0 Introduction 1
1.1 Background 1
1.2 Effectiveness Assessment 1
1.3 Purpose 2
1.4 Rationale for Accomplishing Effectiveness Assessment 2
1.5 Organization of Report 3

2.0 Effectiveness Assessment strategies For Best Management practices 3
2.1 Non-Structural Controls 3
2.2 Structural Controls 4
2.3 Measures of Effectiveness 5
2.3.1 Water Quality Benchmarks 5
2.3.2 Percent Reduction 7
2.3.3 Changes With Time 7

3.0 Qualitative and Semi-quantitative Strategies for Assessing BMP effectiveness 9
3.1 Qualitative Baseline/After Sampling Strategies 9
3.2 Summary of Qualitative and Semi-Quantitative Strategies 10

4.0 Quantitative Designs for assessing bmp effectiveness 13
4.1 Upstream/Downstream Sampling Strategies 13
4.2 Quantitative Baseline/After Sampling Strategies 16

APPENDIX A - Example Implementation Plans
A.1 Hazardous Waste Collection A-1
A.2 Cleaning Streets A-1
A.3 Upstream/Downstream Sampling Design A-2
A.4 Baseline/After Sampling Design A-4
A.5 Retention Systems--Sediment Measures A-5
A.6 Retention Systems--Dissolved and floatable measures A-6

APPENDIX B - Statistics
B.1 Mean B-1
B.2 Variance B-1
B.3 Standard Deviation B-2
B.4 95% Confidence Intervals B-2
B.5 An Example of Calculating Mean, Standard Deviation and Confidence Bounds B-4
B.6 95% Prediction Intervals B-4
B.7 An Example of Calculating Prediction Bounds B-5
B.8 Log transformation B-6
B.9 Percent Reduction B-7
B.10 Simple Linear Regression B-7
B.11 Number of Replicates B-8

1.0 INTRODUCTION

1.1 Background

The County of San Diego initiated Project Clean Water in 2000. The goal of this project is to establish a framework and local commitment for restoring and enhancing the quality of coastal and inland water in the San Diego region. The framework is being built through two levels of technical input. The first level consists of Technical Advisory Committees that provide general guidance and strategic planning on:

- Comprehensive planning
- Legislative and regulatory issues
- Education and resource development and
- Science and technology.

The second level consists of Focused Technical Workgroups that conduct detailed assessments of specific program issues and develop management programs to address them. Different Workgroups are planned over two phases. The first phase Workgroups are:

- Existing Industrial
- Existing Residential
- Existing Municipal
- Land use (including Standard Urban Storm Water Management Programs [SUSMPs])
- Construction and
- Illegal Connection/Illegal Discharge (IC/ID).

The main goal of the Focused Technical Workgroups is to develop Model Urban Runoff Management Plans (URMPs) for the Co-permittees of San Diego County. Each Co-permittee can then draw upon the model plans and customize them to produce Jurisdictional URMPs for its own jurisdiction.

Local commitment has been strongly encouraged. Both the Technical Advisory Committees and Focused Technical Workgroups, although led by San Diego County Employees, are made up of contingents from many Co-permittees, including the City of San Diego, environmental consultants, and industrial representatives. Members of the San Diego Regional Water Quality Control Board (SDRWQCB) have also attended meetings and acted as a sounding board for ideas.

1.2 Effectiveness Assessment

Most models (URMP components) under development in Focused Technical Workgroups will have similar subcomponents that will address:

- Prioritizing pollutants, activities and areas of concern
- Identifying Best Management Practices (BMPs), Best Conventional Technology (BCTs), and Best Available Technology (BAT)
- Suggesting an implementation strategy to reduce or minimize the pollutants, activities, and areas of concern, and
- Providing an approach to an assessment of the effectiveness of BMPs in the program.

Priorities, applicable practices and technologies, and implementation strategies will differ among different jurisdictional URMPs. However, a similar approach can be used to assess the effectiveness of each individual URMP. Indeed, a standardized approach is desirable to facilitate comparisons on a local and regional basis.

Because there is overlap among pollution-reducing activities (BMPs, BCTs and BATs), the assessment of program effectiveness has been organized and presented by pollution-reducing activity. For the purpose of this document, BMPs, BCTs and BATs will all be referred to as BMPs. The rest of this document describes methods of categorizing BMPs and the assessment techniques for each category of BMP.

1.3 Purpose

The purpose of this document is to provide the paradigm for assessing the effectiveness of BMPs, specifically whether there has been a reduction in the pollutant(s) targeted by the BMP. The success of jurisdictional URMP programs, which will target the reduction of a number of pollutants in the Municipal Separate Storm Sewer Systems (MS4) and which will encompass a number of BMPs, will be best assessed by the wet and dry weather monitoring programs of the Copermittees.

1.4 Rationale for Accomplishing Effectiveness Assessment

There are five major reasons for committing resources to effectiveness assessments:

- Demonstration that significant efforts are being made to achieve water quality benchmarks
- Demonstration that Maximum Extent Practicable (MEPs) levels of pollutant reductions are being met
- Rebuttal to claims of polluting by regulators or environmental interveners
- To prevent punitive steps by demonstrating improvement when water quality measures do not yet meet benchmarks and
- To prevent regulatory fines.

In addition, if total maximum daily load reductions (TMDLs) are required in the future, the record of cleanup or pollution reduction in the past and present, if well documented, might be used as offset credit against future required reductions.

1.5 Organization of Report

Section 2 discusses the categories of BMPs. Sampling and analytical designs used to assess effectiveness for a major category of BMPs called non-structural BMPs are presented in Section 3. Section 4 presents strategies for assessing the effectiveness of BMPs in the other major category: structural BMPs. Appendix A details example implementation plans for commonly employed BMPs. Finally, Appendix B contains the mathematical and statistical formulae associated with the presented analytical designs.

2.0 EFFECTIVENESS ASSESSMENT STRATEGIES FOR BEST MANAGEMENT PRACTICES

Best Management Practices, or BMPs, are engineering and non-engineering practices and techniques which mitigate the adverse impact of flooding and surface water quality degradation resulting from land development and urbanization. Engineered BMPs are typically referred to as "structural controls." Engineering practices include the construction of extended detention ponds, infiltration trenches, porous pavements, water quality inlets/outlets, grass swales, filter strips, pollutant/nutrient uptake wetlands, and erosion and sediment controls.

Non-engineering practices include working with policies, regulations, construction and maintenance plans, and public education. These non-engineering practices are often referred to as non-structural controls. Non-structural work often includes the review of stormwater management plans, which ensure resource protection, and ease of obtaining permitting.

The following sections provide a discussion of the important characteristics of non-structural and structural BMPs as they pertain to developing effectiveness assessment strategies.

2.1 Non-Structural Controls

Non-structural Controls are activities aimed at reducing the amount of pollutants that are available for capture and incorporation into storm water runoff. There are two kinds of Non-structural Controls including:

- Source Controls
- Pollution Prevention Educational Programs.

Non-structural source controls include schedules of activities, prohibitions of practices, maintenance procedures, managerial practices, or operational practices that aim to prevent storm water pollution by reducing the potential for contamination. Pollution Prevention educational programs stress the need for pollutant reduction and provide information that the public and municipal, industrial, and construction employees can utilize to participate in pollutant reduction efforts.

Pollution prevention education programs do not involve structures that are amenable to sampling designs based on direct measurements. Measurements of the success of educational programs must involve a qualitative documentation of the level of effort expended over time. These data can be combined with other semi-quantitative or quantitative data related to the quality of receiving waters in a watershed to allow an agency to determine if their BMP implementation is having a real effect on water quality.

The effectiveness of BMPs for non-structural controls is best accomplished by establishing baseline levels of effort and the concentrations of pollutants in receiving waters before the BMPs are implemented. Establishment of these baseline conditions allows documentation of any reduced concentrations that result from the implementation of a BMP program.

2.2 Structural Controls

The important characteristic of structural controls as they pertain to an assessment of their effectiveness is that the controls are constructed to remove pollutants "in-line". This means that they are constructed as an integral part of the hydrologic flow system that conveys the pollutants. Structural controls that are employed in or near source areas are referred to as *Source Controls*. Source controls prevent pollutants from leaving a source area and entering the MS4 system. Conversely, *Treatment Controls* remove pollutants from waters that are already entrained in stormwater. They typically occur within the MS4 system.

Examples of in-line treatment and source control structural BMPs for model Urban Runoff Management Plans (URMPs) are shown in Table 2-1.

Table 2-1. Example Structural BMPs.

Model URMP	Treatment Controls	Source Controls
Construction	Detention ponds	Hazardous waste management pit
Municipal	Non-emergency stormwater facility repairs and construction	Secondary containment enclosures for tank leak and spill control
Land Use (SUSMPS)	Structural Treatment BMPs that filter, treat, or infiltrate runoff	Secondary containment of potential contaminants by berms, dikes or curbs
Existing Industrial	Storm drain filters	Vehicle equipment washing and steam cleaning capture basins
Existing Residential	Infiltration wetlands or grass parkways	Berms around automobile repair and maintenance areas

The in-line characteristics of structural controls allow direct measurements of BMP effectiveness because they provide easily identified control points for measuring flow and retrieving water quality samples. These measurements allow calculation of a quantitative assessment of effectiveness by comparing spatial and temporal trends in constituent concentrations, or loads, at points upstream and downstream of the BMP control.

2.3 Measures of Effectiveness

The effectiveness of BMP programs can be measured either directly or indirectly. ***Indirect Measures* track the level of effort of program BMP activities.** Examples of indirect measures are numbers of inspectors, numbers of notices of violations, or surveys of responses to community outreach. Indirect Measures do not allow a quantitative evaluation of pollutant reductions; however, data can be collected and presented to show that a jurisdiction is making a strong effort in implementing non-structural BMP programs and activities. Examples include graphs that document an increasing level of effort through time (Figure 1), or graphs that document an increase in the level of awareness related to water quality issues as a result of community outreach programs (Figure 2).

***Direct Measures* refer to actual measurements of pollutants of concern, such as concentrations of chemicals or amounts of material extracted by a BMP.** Effectiveness assessments that involve collection of quantitative water quality information can be evaluated using any of the three following techniques:

- Comparison to Water Quality Benchmark Criteria
- Calculating the percent reduction of pollutants
- Determining the temporal changes in the concentrations of pollutants.

2.3.1 Water Quality Benchmarks

Water Quality Benchmarks are goals that can be used to judge if concentrations of pollutants of concern are at, or near, acceptable levels. These goals are water quality criteria established to maintain the quality of receiving waters. Currently, there is no single comprehensive regulatory list of storm water quality criteria that can be used to judge storm water analytical results. The Science and Technology Committee of The Clean Water Program of The County of San Diego is presently in the process of compiling storm water quality benchmarks from key regulatory sources, such as the California Basins Plan, the California Toxics Rule and the USEPA National Pollutant Discharge Elimination System (NPDES) Storm Water Multi-Sector General Permit for Industrial Activities.

Benchmarks are the highest measure of the standard of effectiveness of BMP programs and can be used as ultimate goals against which to track improvements in pollution control. BMPs with effectiveness measures that can be compared to benchmarks offer the only direct way in which a statement of water quality can be made.

Benchmarks may be difficult to meet in many cases. An awareness of this has resulted in the practice of reducing the concentration of pollutants to the Maximum Extent Practicable (MEPs). MEPs consist of SDRWQCB approved proposals from the municipalities of the combination of all BMPs in the URMPs and activities to support them. These BMPs must be effective in reducing levels of pollutants of concern at a reasonable cost. Thus, demonstrations of effectiveness are necessary to gain a MEP. These demonstrations can be accomplished by tracking the percent reduction of a pollutant due to BMPs, or by tracking changes in the concentration of a pollutant with time.

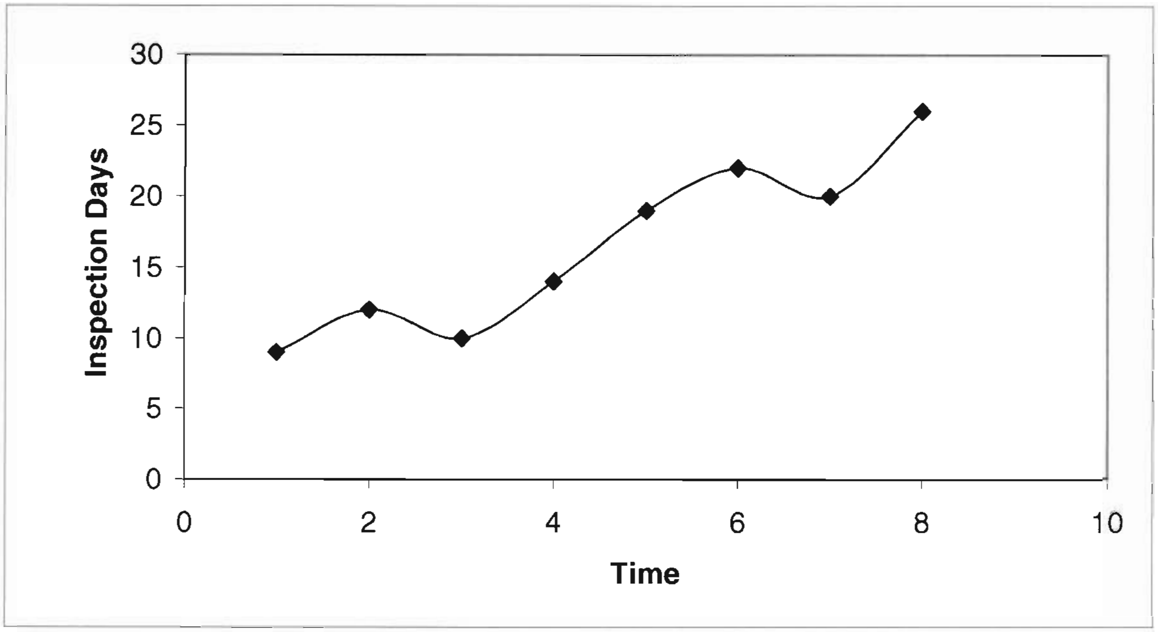

Figure 1. Example plot of increasing effort within a BMP program with time.

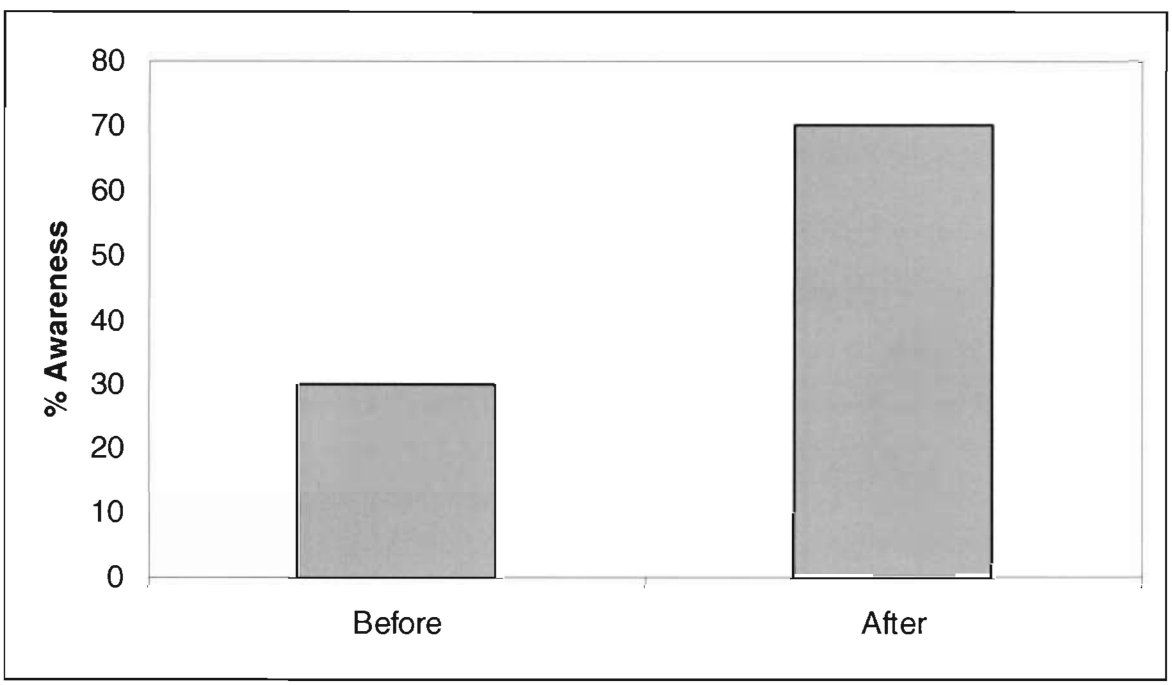

Figure 2. Example plot of water quality awareness issues surveyed before and after implementation of community outreach programs.

2.3.2 Percent Reduction

Percent reduction evaluations measure the change in pollutant concentrations from some previous period, either a baseline or a previous year, and compare those to concentrations after BMP implementation. An effectiveness measure that does not meet a benchmark, or a MEP, may still demonstrate a satisfactory clean-up effort with a large, or statistically significant percent reduction.

2.3.3 Changes With Time

Two primary variables change through time in BMP programs. Those variables include:

- The level of effort expended in the program(s)
- The concentration, or amount, or pollutants targeted by the BMP program(s).

Level of effort tallies may include such things as tallies of expenditures and/or activities related to a BMP program. Combined plots of the level of effort expended in a program and pollutant concentrations, or amounts, can be used to assess the effectiveness of a BMP program on a per unit basis.

For example, the changes in the concentration, or amount of a pollutant may tend to decrease, or level-off, with time (Figure 3a.) even though the level of effort expended remains constant as the program progresses (Figure 3b.). The combined plot showing the amount of pollutant removed per unit of effort expended has peaked (Figure 3c.). If the level of effort is constant, or increasing, while the pollutant reduction benefit is decreasing, an important part of a MEP is demonstrated – that costs associated with increasing the level of effort are unreasonable compared to the pollution control benefits achieved.

Figure 4a. and 4b. show that the amount of pollutant removed during the implementation of a BMP program increases dramatically early in the program. However, as time progresses the program is less effective, even though the level of effort is being expended increases as the program progresses. This situation may indicate that pollutant concentrations are approaching ambient, or background concentrations as the program progresses. Additional levels of effort in that program are not cost-effective. The resulting conclusion that the long-term cost of increasing the level of effort for that particular program is prohibitive compared to the benefit received is apparent in Figure 4.c.

A final tool that can display changes with time involves the use of linear regression models. This tool may show there are statistically significant improvements taking place in a program that are not readily apparent otherwise (e.g. Figure 5). The consistent improvement can demonstrate reasonable intent and effectiveness within a program. Linear regression methodology is discussed in Appendix B.

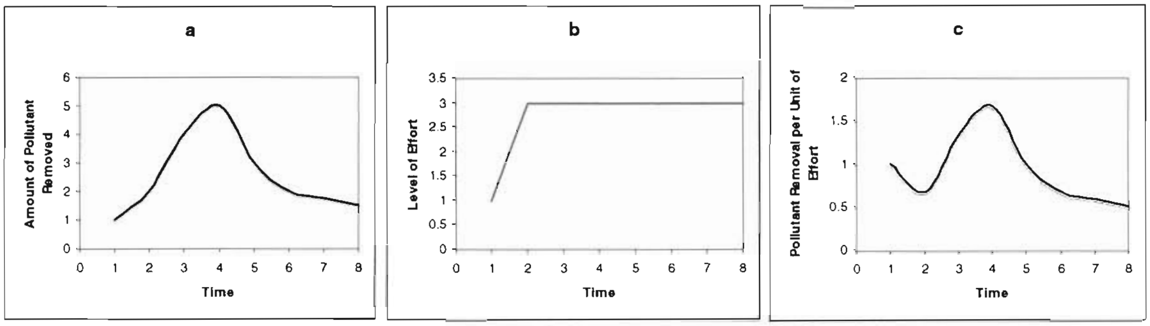

Figure 3. Pollutant removal with constant level of effort.

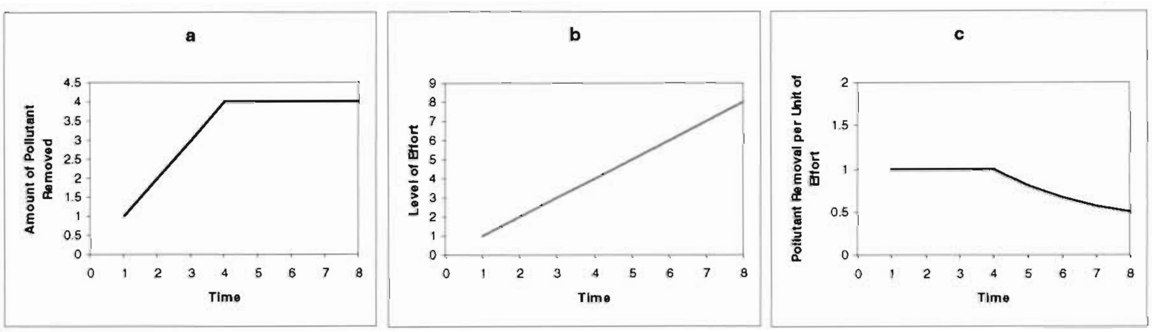

Figure 4. Pollutant removal with increasing level of effort.

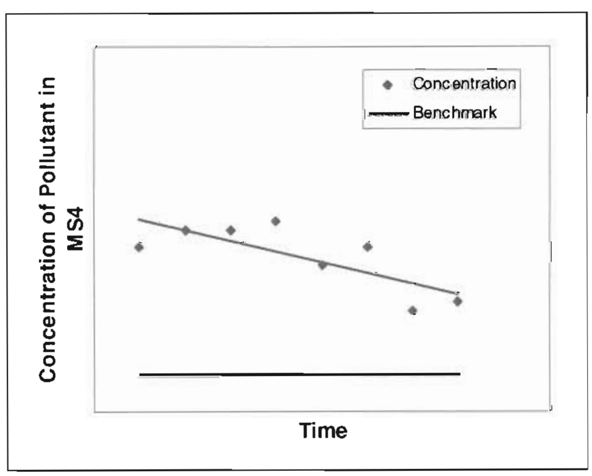

Figure 5. Concentration of pollutant compared to benchmark value.

3.0 QUALITATIVE AND SEMI-QUANTITATIVE STRATEGIES FOR ASSESSING BMP EFFECTIVENESS

Qualitative strategies provide useful visual documentation and records of the level of effort being expended on BMP programs, and are usually easier and less expensive to implement compared to quantitative strategies. Qualitative and semi-quantitative assessments of the effectiveness of BMP programs focus on providing data that can be used to establish if the implementation of non-structural BMPs is having an effect on the overall water quality in a receiving water. These assessments include documenting the amount of effort that is being expended on BMP programs with information such as:

- The number and type of BMPs that are being implemented
- The frequency and duration of BMP implementation.

This type of assessment does not involve sampling or measuring pollutant concentrations in storm water. However, these qualitative measurements provide important information that can be combined with more quantitative data so the relationships between the levels of effort expended on BMP programs and actual pollutant reductions can be evaluated.

Examples of qualitative assessment tools include photographs of all of the necessary BMPs properly installed at a construction site. Photographs of berms or dams around vehicle refueling areas are other examples. Calculations of the volume of material that can be contained by dams and berms are another qualitative measure. Examples of data that can be collected to document and track the level of effort expended in a BMP program include tracking (also called "tallies") the number of public surveys made, and the numbers of inspectors involved in a program.

3.1 Qualitative Baseline/After Sampling Strategies

Baseline/After sampling strategies can provide quantitative or qualitative data. Measurements are taken in two or more time periods including:

1. Baseline measurements before the BMPs are initiated, increased, or changed
2. Later measurements at single or multiple times after BMP program is initiated.

One of the main uses of this design is for the documentation of the level of increased activity in Non-structural Source Control or Pollution Prevention efforts.

Baselines are extremely important for this design because all future comparisons showing improvements will be made relative to the Baseline. If a Baseline is not taken, most of the improvements that often occur at the beginning of a program will not be documented, and credit for them will be lost.

The Baseline/After sampling design is appropriate for BMPs such as cleaning and maintenance BMPs where materials removed can be measured. The design is also suitable for hazardous waste and used oil collection sites where amounts brought in can be documented. Pollution Prevention Activities such as records of corrections of illicit connections, or notices of violation, are also examples of documented activities that can be tracked in time.

3.2 Summary of Qualitative and Semi-Quantitative Strategies

Table 3-1 provides a summary of non-structural BMPs and appropriate effectiveness assessment strategies. There are often several appropriate assessment strategies for one BMP. This is especially the case when there is not a clear distinction between the structural or non-structural nature of the BMP. Table 3-1 also provides a listing of potential monitoring components that can be tracked, a listing of constituents of concern that a BMP may target, and likely data evaluation techniques.

Most of the non-structural BMPs rely on qualitative assessment such as tracking the number of miles of streets that are cleaned on an annual basis. These tallies can be compared in Baseline/After design evaluations, or plotted as a time series to document the level of effort in a BMP program as shown in Figures 1 and 2. Tallies can be documented in units of number of occurrences, number maintained, acres maintained, miles, or other appropriate measures.

The importance of baseline documentation can not be overemphasized. The development of a baseline provides for the widest selection of assessment strategies. For example, in Table 3-1 if "vehicle and equipment cleaning and maintenance" has a baseline, then the percentage improvement from what was done before, or what was always done can be shown (e.g., "120 vehicles now serviced to guard against oil and grease leaks in contrast to 22 vehicles last year"). Without a baseline, the assessment can only track changes that occur after the initiation of a BMP Program.

Table 3-1. Summary of non-structural BMPs and appropriate effectiveness assessment strategies.

BMP Category	Sampling Design				Comparison Standard			Monitoring Components									Constituents of Concern							
Non-structural Source Control BMPs	Upstream/ Downstream	Baseline/After	Qualitative/Effort Documentation (Tally)	Photographs	Benchmarks	Changes Over Time	Percent Change	Miles Cleaned	Miles Mult. By Number of Times	Miles Repaired	New Miles or Acres	Number Maintained	Number Repaired	Number Replaced	Number of Uses (Tally)	Photographs	Sediment	Nutrients	Heavy Metals	Toxic Materials	Floatable Materials	Oxygen Demanding Substances	Oil and Grease	Bacteria & Viruses
Maintenance Practices																								
Catch basin insert cleaning		x	x			x	x					x	x				x		x		x	x	x	
Street and parking lot cleaning		x	x			x	x	x	x			x					x		x		x		x	
Road repair		x	x			x	x			x									x		x		x	
Ditch cleaning, stabilization, and erosion control		x	x			x	x	x					x			x	x	x			x	x		
Sediment and floatables removal from constructed BMPs		x	x			x	x					x	x				x				x			
Vegetation maintenance in constructed BMPs		x	x			x	x					x					x	x			x	x		
Repair of outlet structures		x	x			x	x						x	x			x							
Replacement of filtration media	x	x	x			x	x					x		x			x				x		x	
Good Housekeeping/Management Practices																								
Hazardous materials and waste management		x	x	x		x	x					x			x	x	x	x	x	x		x	x	
Vehicle and equipment cleaning and maintenance		x	x			x	x					x	x							x			x	
Spill prevention		x	x			x	x									x			x	x	x		x	
Stream crossing or diversion		x	x	x		x						x			x	x	x						x	
Dry cleanup technique		x	x			x											x		x	x	x			
Water conservation		x	x			x	x								x			x		x				
Spill cleanup		x	x			x	x								x				x	x		x	x	
Minimizing pesticide, herbicide & fertilizer use		x	x			x	x								x			x				x		
Illicit Discharge Detection and Elimination		x	x			x	x			x			x		x	x	x	x	x	x	x	x	x	x
Irrigation control		x	x			x					x	x	x			x	x	x		x		x		

Table 3-1. Continued.

BMP Category: *Non-structural Source Control BMPs*	Sampling Design: Upstream/ Downstream	Sampling Design: Baseline/After	Sampling Design: Qualitative/Effort Documentation (Tally)	Sampling Design: Photographs	Comparison Standard: Benchmarks	Comparison Standard: Changes Over Time	Comparison Standard: Percent Change	Monitoring Components: Miles Cleaned	Monitoring Components: Miles Mult. By Number of Times	Monitoring Components: Miles Repaired	Monitoring Components: New Miles or Acres	Monitoring Components: Number Maintained	Monitoring Components: Number Repaired	Monitoring Components: Number Replaced	Monitoring Components: Number of Uses (Tally)	Monitoring Components: Photographs	Constituents of Concern: Sediment	Constituents of Concern: Nutrients	Constituents of Concern: Heavy Metals	Constituents of Concern: Toxic Materials	Constituents of Concern: Floatable Materials	Constituents of Concern: Oxygen Demanding Substances	Constituents of Concern: Oil and Grease	Constituents of Concern: Bacteria & Viruses
Minimize Soil and Vegetation Disturbance																								
Maintain naturally vegetated areas to minimize runoff		x	x			x	x				x	x				x	x	x						
Minimize active construction areas (fingerprinting) to minimize erosion		x	x												x	x	x	x						
Construction activity scheduling to minimize soil exposure		x	x												x		x							
Sensitive Area Protection		x	x			x	x								x	x	x	x	x	x	x	x	x	x
Minimize Impervious Surfaces																								
Porous pavement or modular block pavers		x	x				x		x		x						x		x	x			x	
Reduce street width		x	x				x		x		x						x		x	x			x	
Open Space Preservation		x	x			x	x					x			x		x	x	x	x	x	x	x	x
Public Education/Community Outreach																								
Storm drain inlet stenceling		x	x									x			x		x	x	x	x	x	x	x	x
Proper methods of application of, and alternatives to, pesticides/herbicides		x	x												x					x				
Proper application of fertilizers		x	x			x	x								x			x				x		
Household hazardous materials disposal		x	x			x	x								x			x				x		
Pet waste disposal		x	x			x		x			x				x							x		x
Provide materials via flyers, newsletters, workshops,		x	x			x	x								x		x	x	x	x	x	x	x	x
Employee Training		x	x			x	x								x		x	x	x	x	x	x	x	x
Recycling Facilities		x	x			x	x					x	x		x				x	x	x		x	

4.0 QUANTITATIVE DESIGNS FOR ASSESSING BMP EFFECTIVENESS

There are two useful strategies for quantitatively measuring the effectiveness of BMPs including Baseline/After strategies previously discussed under qualitative assessments, and Upstream/Downstream sampling. The Upstream/Downstream sampling strategies are used primarily for Treatment Control BMPs. The quantitative Baseline/After sampling designs are used primarily for Source Control and Pollution Prevention BMPs. These strategies are discussed in detail in the following subsections and are listed in Table 4-1. Table 4-1 also provides a listing of constituents of concern that a BMP may target, and likely data evaluation techniques.

4.1 Upstream/Downstream Sampling Strategies

Upstream/Downstream sampling designs result in the most quantifiable of all effectiveness assessments. Upstream/Downstream designs have nearly simultaneous measurements of pollutant concentrations or load at two or more sites relative to a BMP structure. At a minimum, one sampling site is upstream of the BMP structure and the other is downstream. Water samples are collected and the concentrations of pollutants are directly measured. The downstream minus upstream difference in pollutant level is a measure of the amount of pollutants removed. The amount removed can be expressed in several ways, including:

- Changes in concentration between upstream and downstream sampling locations
- The percent of pollutant removed between sampling stations
- Comparisons of downstream pollutant concentrations to water quality benchmark objectives
- Temporal analyses that indicate year-to-year reduction of pollutants.

The advantage of this design is that quantitative data is produced that directly demonstrates pollutant reduction from the BMP. The Upstream/Downstream design is only appropriate for in-line Structural Treatment Control BMPs with well-defined water entry and exit points. Examples include flow-through detention ponds for sediment deposition, or screens for removing oils and other liquid and solid floatables. Structural Source Control BMPs, such as a sediment fence or a containment enclosure, are a special case of this design because barriers prevent flow through the source from upstream sources. In cases like this, upstream concentrations or loads are assumed to equal zero, while the downstream measure is the amount or concentration of the material that gets past the barrier (Table 4-1b).

Table 4-1. Summary of (a) structural treatment controls and (b) structural source controls with appropriate effectiveness assessment strategies.

BMP Category	Sampling Design				Comparison Standard			Constituents of Concern							
Structural Treatment Controls	Upstream/ Downstream	Baseline/After	Qualitative/Effort Documentation (Tally)	Photographs	Benchmarks	Changes Over Time	Percent Change	Sediment	Nutrients	Heavy Metals	Toxic Materials	Floatable Materials	Oxygen Demanding Substances	Oils and Grease	Bacteria and Viruses
Filtration Systems															
Surface sand filter	x		x		x	x	x	x							
Underground vault sand filter	x		x		x	x	x	x							
Compound filter	x		x		x	x	x	x				x		x	
Filter fence	x		x		x	x	x	x				x			
Biofiltration	x		x		x	x	x								x
Storm drain filters	x		x		x	x	x					x			
Curb inlet filters	x		x		x	x	x					x			
Gravel bag berm	x		x		x	x	x	x				x			
Separator Systems															
Sediment	x				x	x	x	x				x			
Oil and grease	x				x	x	x							x	
Screening: flotation	x				x	x	x					x			
Settling: centrifuge	x				x	x	x	x				x			
Injector /Light systems															
Ozone	x				x	x	x								x
Ultra-violet	x				x	x	x								x
Infiltration Systems															
Basins and channels		x	x			x	x	x		x	x	x		x	
Porous pavement		x	x			x	x	x							
Trenches and walls		x	x			x	x	x				x			
Vegetated Systems															
Grass filter strips		x	x			x	x	x	x	x	x	x	x	x	x
Grassy swales		x	x			x	x	x	x	x	x	x	x	x	x

Table 4-1. Continued.

BMP Category	Sampling Design				Comparison Standard			Constituents of Concern							
Structural Source Controls	Upstream/ Downstream	Baseline/After	Qualitative/Effort Documentation (Tally)	Photographs	Benchmarks	Changes Over Time	Percent Change	Sediment	Nutrients	Heavy Metals	Toxic Materials	Floatable Materials	Oxygen Demanding Substances	Oils and Grease	Bacteria and Viruses
Retention Systems															
Catch basin	x		x		x	x	x	x	x			x	x		
Constructed Wetlands Systems	x	x			x	x	x	x	x	x	x	x	x		x
Ponds		x	x			x	x	x	x	x	x	x	x		x
Tanks		x	x			x	x	x	x	x	x	x	x		x
Tunnels		x	x			x	x	x	x			x	x		x
Vaults and pipes		x	x			x	x	x	x	x	x	x	x		x
Bioretention		x	x			x	x	x	x	x	x	x	x		x
Check dam		x	x			x	x	x	x	x	x	x	x		x
Berms		x	x			x	x	x	x	x	x	x	x		x
Slope Protection and Stabilization															
Retaining walls		x*	x	x				x							
Fabrics and blankets		x	x	x				x							
Jute		x	x	x				x							
Geotextiles		x	x	x				x							
Geogrids		x	x	x				x							
Mulch		x	x	x				x							
Seeding		x	x	x				x							
Riprap		x	x	x				x							
Channelization		x	x	x				x							
Soil binders		x	x	x				x							
Earth dikes		x	x					x							
Drainage swales and ditches		x	x	x				x							

* Special case of design where upstream load is assumed to be zero and downstream measure is the current or concentration of material that gets by the barrier.

4.2 Quantitative Baseline/After Sampling Strategies

There are many instances where quantitative Baseline/After sampling designs are appropriate for structural and non-structural BMPs. Examples of structural BMPs where Baseline/After sampling designs are appropriate include infiltration systems, wet ponds, or retention basins because they are not in-line but end-of-line. Establishing chemical conditions at the base of a watershed and tracking any changes in those concentrations through time provides data to quantify the cumulative effect of structural and non-structural BMP programs as they are implemented.

APPENDIX A

Example Implementation Plans

IMPLEMENTATION OF EFFECTIVENESS ASSESSMENTS

This Appendix contains example implementation plans for BMPs. The organization is by sampling designs that were presented in Sections 3.2, 4.1, and 4.2. Tables 3-1 and 4-1 present lists of BMPs and the appropriate sampling design(s) for assessing effectiveness. Persons who have already selected a BMP can look up the BMP in Table 3-1 (Non-structural) or Table 4-1 (Structural) and find the sampling design appropriate to its assessment, and then refer to the appropriate sampling design descriptions below for an appropriate effectiveness assessment approach.

A.1 Hazardous Waste Collection

Daily records of counts of each category of pollutant brought to collection sites are accumulated and presented as monthly totals throughout the year and as annual totals for tracking over time. Example categories are paint, pesticides, and oils. The amounts of pollutants prevented from entering the MS4 system annually are estimated from determining the weights of waste for each count category, and the pollutants per waste category from waste category specifications or regulatory agency literature. The annual amount of any pollutant is estimated by multiplying the count by the average weight of waste per count and by the concentration of pollutant per unit weight for each category of hazardous waste and summed over all hazardous waste categories. Annual amounts should be compared to the baseline to document increases or decreases of collection quantities.

A.2 Cleaning Streets

The following method is adapted from sequential quality control analyses. It provides the most cost effective method (least number of samples) of determining whether or not a baseline has been exceeded or improvement has occurred.

The Baseline consists of measuring the weight of material and concentration of pollutants in the material at the present time. Thereafter, the continued program of measurements contributes to the 'After' period.

There are three categories of pollutants found in street cleaning: sediments, chemical pollutants, and trash. These should all be sampled from the street cleaners and treated as described in the previous sections.

Initially, the volume, weight and chemical composition of 10 street-sweeper loads are measured with respect to sediments, chemical pollutants and trash. From this information, means and prediction bounds (as described in Appendix B) are set up in the form of control charts (Appendix B, Figure 1). Thereafter, one street-sweeper a week is selected randomly and its load measured. This is repeated four times, for one month. If these loads fall within the prediction bounds established, the program can shift to measuring one street-sweeper per month.

If the sediments, chemical pollutants or trash fall outside the prediction bounds of the control charts, the program shifts back to one street-sweeper load per week for four weeks. At this point, a determination is made as to whether the outside measurement was an aberration or a change has taken place as indicated by three of the four subsequent measurements also falling outside the prediction interval on the same side.

If the outlier was an aberration, the sampling frequency reverts to once a month and recalculation of the prediction bounds is made using all the data of the calendar year.

If the outlier and subsequent measurements were below the lower bound, the program reverts to the once a month sampling and the prediction interval should be recalculated at the end of the rainfall year.

If the outlier and subsequent measurements were above the upper bound, then efforts should be undertaken to identify the source of the increased pollution and new BMPs might be considered. Additionally, the prediction bounds should be recalculated immediately using the last five measurements and the subsequent five. These subsequent samples should be taken weekly. Then monthly sampling can be resumed.

After two years (a minimum of 36 samples), the relationships between sweeper load volume and each of the pollutants (sediments, chemical pollutants, and trash) should be determined by a statistician. The information can be useful in making a determination of whether it is sufficient to estimate street sweeping pollutant loads from sweeper load volumes rather than direct sampling.

A.3 Upstream/Downstream Sampling Design

The Upstream/Downstream design is most appropriate for in-line Structural Treatment Control BMPs and Structural Source Control BMPs. Structural Source Control BMPs are a special case because no upstream measurements can be taken and, therefore, the upstream measurements are assumed to equal zero when performing the calculations.

The following outlines a model approach for the Upstream/Downstream design:

- Take a sample from waters entering the Structural Treatment Control (no sample is necessarily taken for source control, unless knowledge of the amount of material or contaminate being contained or controlled in desired).
- If there is more than one incoming stream, obtain samples from each stream.
- Simultaneously, or with proper lag time, sample downstream of the Structural Treatment Control or the Structural Source Control. Each upstream measurement is considered paired to its simultaneous or possibly lagged downstream counterpart (if timing samples is too difficult, take samples at random times upstream and downstream).
- If there is more than one exit, sample at all exits.
- Repeat this approximately ten (10) times during the first storm. These repetitions are called replicates.
- Once samples from the first storm are measured, the mean, variance and reduction of a pollutant of interest can be calculated and this information can be used through power analyses to determine the most cost effective number of repetitions during a storm. Depending on the variance of that pollutant in the replicate samples taken from the first

storm, the number of future replicates needed to demonstrate the expected decrease can be looked up in Appendix B.

- Sample two high rainfall events to demonstrate efficacy under higher BMP stress conditions and one low rainfall event to see that the BMP also works under low flow conditions. Individual BMPs also need to be measured more frequently, at least initially, to demonstrate efficacy, and to learn and demonstrate the effectiveness of maintenance schedules.
- Some BMPs can be operated at different levels of treatment. If the level of treatment by the BMP can be changed, the above steps should be repeated at each level, or at least the minimum level of treatment, to demonstrate its effectiveness.
- The effectiveness assessments are summarized in a presentation of the following for every constituent measured (See Appendix B for details of calculations):
  - Mean for each storm
  - Percent reduction from upstream, if available, for each storm
  - Mean of the individual storm means (grand mean) for the baseline year, and
  - Percent reductions from the baseline grand mean.
- Plots over time of all above measures and levels of effort of BMP activity. These plots will have benchmarks noted on plots of absolute values and baselines noted on plots of reductions (e.g., Figure A-1).

Temporal tracking for individual BMPs is important to:

- Demonstrate initial efficacy (the BMP is new) under different load conditions,
- Indicate maintenance is required as efficacy measures drop off. (For this indication, upstream/downstream comparisons are needed to validate that loss of efficacy downstream is due to BMP and not to increased pollutant levels upstream.)

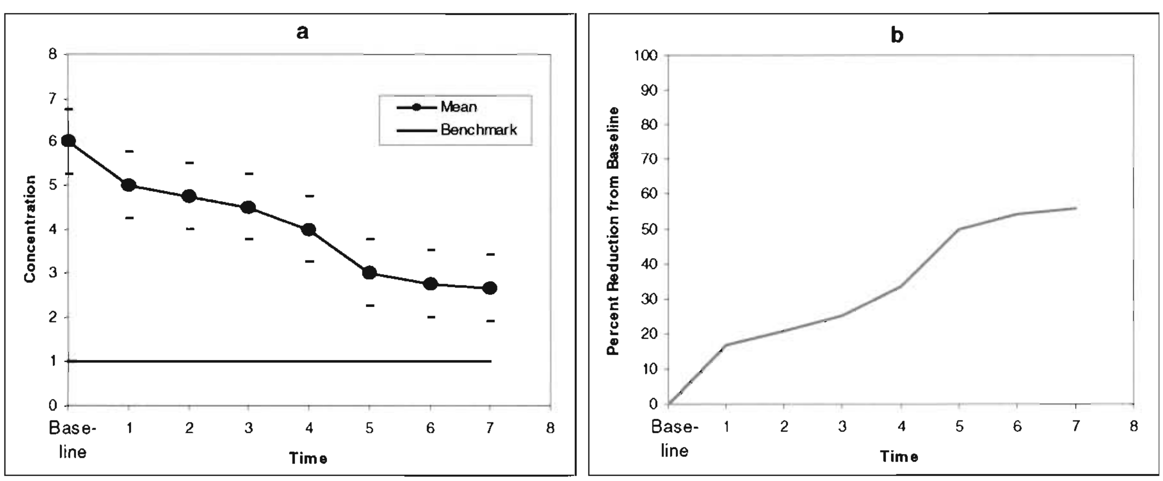

Figure A-1. Effective assessment example plot.

Temporal tracking for Upstream/Downstream in MS4 or receiving waters for multiple BMPs or, for example, a municipal BMP such as a detention basin, is important to:

- Demonstrate that multiple storms per year show efficacy under different rainfall or storm-flow conditions
- Detect increases in pollutant levels upstream indicating additional sources
- Demonstrate reductions in levels of pollutants of concern that are moving towards, but may have not yet reached, benchmarks. These reductions are not always reflected in downstream values alone because they can increase in the face of increasing upstream values. Consequently, this demonstration is best measured by downstream minus upstream differences.

In most cases for chemical concentration data it is appropriate to mathematically transform data measurements to meet validity requirements for statistical analysis. Appendix B describes how to do the transformation and how to invert the transformation for summary results.

A.4 Baseline/After Sampling Design

The Baseline/After design is used for Structural Source Control BMPs that retain received waters or prevent pollutants from leaving their source (e.g., trash enclosures), or for Non-structural Source Control BMPs, that involve cleaning or maintenance. The purpose of these designs is to measure the quantity of pollutants not being released to storm waters.

Comparisons to the benchmarks cannot be made as they can for Upstream/Downstream designs. Nor can percent reductions in storm waters be calculated directly. Values can be tracked over time and annual percent change over time can be calculated. Note, these calculations should be accompanied with level of effort numbers to demonstrate the cost effectiveness of a clean-up effort.

The sampling design has three components: When to measure (during the Baseline period and After periods), what to measure, and how to present the results.

Baseline measurements are taken before the BMP is functioning or at the beginning of the season as a measure of the initial state of the system.

It is imperative to measure a Baseline as a representation of the 'Before' condition, i.e., the state of pollution reduction and clean-up effort before an increased level of BMP effort is begun or new BMPs are brought on line.

Sampling frequency in a Baseline period can be different than in an After period. In the Baseline year, samples should be taken at every opportunity, that is, every time a cleaning effort is made. For BMP structures that are cleaned less than four times per year, every effort should be made to go back and recover past data on volume, mass, and concentrations of material removed. Calculation of means and standard deviations for the Baseline can be used to calculate the number of times per year the BMP should be sampled in the After period to achieve a percent level of precision to detect pollution reduction.

In the After period, when to measure is, in general, determined by the cleaning schedule. For example, oil & grease filters are weighed when they are changed, street cleaning trucks are weighed when they return with a load, catchment basins are sampled when they are cleaned once per year just before the wet weather season.

In general, for all the Structural and Non-structural Source Control BMPs, pollutant material or materials containing pollutants are collected and ultimately disposed of, usually in a landfill. In most cases, the amount of collected substance is estimated and the concentration of pollutants within that substance is determined.

Specific requirements that depend on the type of BMP are discussed in the following subsections. Infiltration and retention ponds are Structural BMPs designed to capture storm waters and collect sediment, particulates, and floatables. Street cleaning and storm drain maintenance are Non-Structural BMPs that remove sediment and floatables left behind. The Baseline/After design also applies to BMPs like hazardous waste disposal and alternative products usage where documentation of materials collected or sold can be maintained.

The effectiveness assessments are summarized in the following presentation for various constituents measured (See Appendix B for details of calculations):

- Mean of all samples for the baseline year
- Percent reductions from the baseline mean
- Plots over time of all above measures and levels of effort of BMP activity. These plots will have benchmarks noted on plots of absolute values and baselines noted on plots of reductions.

A.5 Retention Systems--Sediment Measures

Sediment measurements are made in catchment basins, retention ponds, infiltration systems, and other water collection areas without storm water throughput. For purposes of this discussion, we will call these retention systems.

Different retention systems are cleaned on different frequencies. For example, infiltration basins are restored to their original state very infrequently, whereas, catchment basins are sometimes cleaned more than once a year. In general, measurements should be made on retention systems when they are cleaned. An exception is if measurements of ***rates*** of retention are required, more frequent sample times may be needed. The description of the sampling program below is presented on an annual basis but may be applied more or less frequently, as appropriate.

Initially from engineering drawings when built, or as part of the Baseline, total potential sediment volume (that the system can hold when it is full) of any retention system is estimated. At the beginning of the storm season, an initial volume of material in the retention system, if any, is measured or estimated from the shape and size of the system.

When the material is cleaned out, samples of the material are taken and the weights of different size classes and concentrations of sediment pollutants are measured in the laboratory. Sediment classes should be divided into clays, muds, silts, fine sands, coarse sands, fine gravels, and coarse gravels to ensure consistency among measures from different BMPs and interpretability with respect to how these sediments potentially affect the environment.

Samples must be taken from the total amount removed to obtain the size class and pollutant estimates. These samples can either be taken in situ before removal, or from the early, middle and late truckloads of material as they are being removed. In either case, there will be a number of samples with the weight of material in each size class. The total sediment in the system in each size class is estimated from the samples using the formulae in Appendix B. Sediment pollutants are analyzed from whole sediment samples.

After minus initial differences are calculated for each cleaning. This is a measure of the sediment captured by the system.

A.6 Retention Systems--Dissolved and floatable measures

Generally, dissolved pollutants and oils and greases are not measured in the waters in retention systems because, ultimately, they join with the sediment as sediment pollutants. These are dealt with above. Floatable items usually have an insignificant Baseline, meaning the Baseline can be assumed to be zero. What is important is directly measuring and tracking the amount removed per year. This is done by measuring volumes or by weighing the material removed. Sometimes it is useful to classify the material so that other Non-structural Pollution Prevention BMPs can be initiated or intensified.

APPENDIX B

Statistics

B.1 Mean

The mean is an estimate of the 'center' of the sample data. It should be approximately the same size as the data, larger than the smallest number and smaller than the largest number. A mean, or an average, is calculated by summing together all the data and then dividing the sum by the number of measurements summed. The formula for this calculation is:

$$Mean = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

where

x_i = the i^{th} measurement

n = the number of measurements

$\sum_{i=1}^{n} x_i =$ the sum of measurements from the first (i=1) through the last (i=n).

Note, an easy way to calculate the mean is to use functions in spreadsheet programs.

B.2 Variance

The variance, or s^2, is a measure of the spread of the sampling data. The larger the spread, the larger s^2. The variance is calculated by summing the squares of the sampling data, subtracting from this sum n times the mean squared, and then dividing the difference by one less than the number of observations. The variance is never negative; a negative value is an indication of an arithmetic error. The formula for the variance is

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

where

x_i^2 = the i^{th} measurement squared

n = the number of measurements

$\sum_{i=1}^{n} x_i^2$ = the sum of measurements squared from the first (i=1) through the last (i=n).

$\bar{x}$ = the mean.

Note, the units of the variance are the units squared of the measurements. To return to the original unit, the standard deviation is calculated.

B.3 Standard Deviation

The standard deviation is the square root of the variance.

$$s = \sqrt{s^2}$$

where

s = the standard deviation
s^2 = the variance.

Note, the easiest way to calculate a standard deviation is to use functions in spreadsheet programs.

B.4 95% Confidence Intervals

A 95% Confidence interval is stated as $\bar{x}$ ± Y, where $\bar{x}$ is the mean, and Y is the product of a constant, T, times the standard deviation, s, times the square root of one over the number of observations, $\sqrt{1/n}$. A 95% confidence interval means that you are 95% certain that the true mean amount of pollutant lies in this interval. The formula for the confidence interval is

$$\bar{x} \pm T * s * \sqrt{\frac{1}{n}}$$

Where

$\bar{x}$ = the mean
s = the standard deviation
n = the number of measurements
T is a constant from Table B.1.

A 95% confidence interval can also be presented as limits by carrying out the implied arithmetic to arrive at the minimum and maximum, as follows:

$$\left[\bar{x} - T * s * \sqrt{\frac{1}{n}}, \bar{x} + T * s * \sqrt{\frac{1}{n}}\right].$$

Table B.1 presents two-sided 95% t-values. The constant T is found by going down rows until the sample size is reached. If the sample size is not one of the values in the table, use the next *lower* sample size value, which gives a higher T. For example, T=2.365 for a sample of 8 measurements. If there are 40 measurements, then T=2.042, from the next smaller sample size. A confidence interval cannot be calculated from one observation; it takes at least two observations to calculate a standard deviation. Thus, Table B.1 begins with a sample size of 2.

Table B.1. T constants

Sample Size	T
2	12.706
3	4.303
4	2.182
5	2.776
6	2.571
7	2.447
8	2.365
9	2.306
10	2.262
11	2.228
12	2.201
13	2.179
14	2.160
15	2.145
16	2.131
17	2.120
18	2.110
19	2.101
20	2.093
21	2.086
22	2.080
23	2.074
24	2.069
25	2.064
26	2.060
27	2.056
28	2.052
29	2.048
30	2.045
31	2.042
41	2.021
61	2.000
121	1.980
>121	1.960

B.5 An Example of Calculating Mean, Standard Deviation and Confidence Bounds

Let's assume you've collected the following five measurements:

3.5, 6.8, 4.9, 2.3, 8.4.

Then,

n=5
$\bar{x}$ = 5.180
s = 2.459.

The 95% confidence interval is

$$5.180 \pm 2.776 * 2.459 * \sqrt{\frac{1}{5}} =$$

5.18 ± 3.053 =

[2.127, 8.233].

B.6 95% Prediction Intervals

A 95% Prediction interval is stated as $\bar{x}$ ± Y, where $\bar{x}$ is the mean, and Y is the product of a constant, T, times the standard deviation, s, times the square root of one plus one over the number of observations. A 95% prediction interval means that you are 95% certain that the next single observation on the amount of pollutant will lie in this interval. The formula for the prediction interval is

$$\bar{x} \pm T * s * \sqrt{1 + \frac{1}{n}}$$

Where

$\bar{x}$ = the mean
s = the standard deviation
n = the number of measurements
T is a constant from Table B.1 in Section B.4.

A 95% prediction interval can also be presented as limits by carrying out the implied arithmetic to arrive at the minimum and maximum, as follows:

$$\left[\bar{x} - T * s * \sqrt{1 + \frac{1}{n}}, \bar{x} + T * s * \sqrt{1 + \frac{1}{n}}\right].$$

Prediction interval calculations differ from those of confidence intervals in the additional one inside the square root sign. Prediction intervals are always wider than confidence intervals. This reflects the fact that individual observations are more variable than means.

B.7 An Example of Calculating Prediction Bounds.

As in Section B.5, let's assume you've collected the following five measurements:

3.5, 6.8, 4.9, 2.3, 8.4.

Then,

n=5
$\bar{x}$ = 5.180
s = 2.459.

The 95% prediction interval is

$$5.180 \pm 2.776 * 2.459 * \sqrt{1 + \frac{1}{5}} =$$

$$5.18 \pm 7.48 =$$

[-2.30, 12.66].

Note that this interval is much wider than the 95% confidence interval in Section B.5.

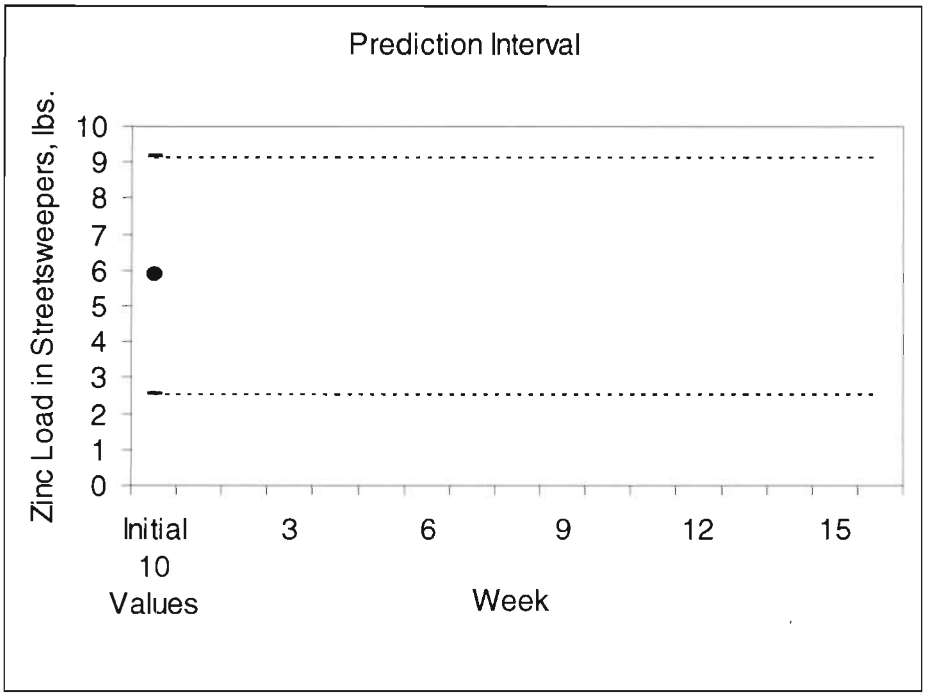

Figure B-1. Example of prediction interval plot.

B.8 Log transformation

Most of the constituents analyzed for in regards to measuring BMP effectiveness require log transformation for analysis. Either natural logarithms or log base 10 may be used.

For the Upstream/Downstream or Before/After designs where differences will be taken, log transformations must be performed before subtracting.

After log transforming, then means, standard deviations, and confidence bounds may be calculated.

It is usually desirable to inverse transform the means and the confidence bounds to return to the original units. This is accomplished by raising 10 (or 'e' for natural logs) to a power equal to the mean or the confidence limits. For example, if the mean and confidence limits on log transformed measurements were the following:

Mean = 2.75
Confidence interval=2.75 ±1.53
Lower limit = 2.75 – 1.53 = 1.22
Upper limit = 2.75 + 1.53 = 4.28

$10^{2.75} = 562.34$
$10^{1.22} = 16.60$
$10^{4.28} = 19{,}054.61$

As this example shows, after inverse transforming the mean and confidence limits, the confidence limits are no longer symmetrical around the mean; i.e., 562.34-16.60 ≠ 19064.61 – 562.34.

If the analyses were performed on differences, then 100 times (10 (or 'e' for natural logs) to a power equal to the mean or the confidence limits) minus 1 equals the percent change from upstream to downstream. For example, if the

mean difference = 0.21
$100*(10^{0.21} - 1) = 100*(0.6166 - 1) = 100*(-0.3834) = -38.34\%$

or, an average 38.34% decrease.

B.9 Percent Reduction

In calculating Percent Reduction, a current vale is compared to a reference value. The reference value is typically the baseline value, last year's value or the first year of the BMP's value. The formula for Percent Reduction is the following:

$$\text{Percent Reduction} = 100 * \frac{\text{Current} - \text{Reference}}{\text{Reference}}$$

where Current is the current value and Reference is the reference value.

B.10 Simple Linear Regression

Simple linear regression is a statistical technique for fitting a straight line to sampling data. For BMPs, the horizontal axis, or X-axis, is elapsed years, from 0 on the left increasing towards the right. The vertical, or Y-axis, is the constituent of interest, log transformed, if necessary. If the BMP has been effective, then the line should be sloped downwards to the right. The slope of the line measures the annual rate of change. If the measurements were log transformed, then $100*(10^{slope} - 1)$ is the average annual percent change.

Simple linear regression should not be calculated by hand. Spreadsheet software can calculate the slope, as well as statistical analysis software. These software packages will also calculate a P-value for testing the hypothesis that the slope is zero (no change per year). The P-value is the probability of getting the measurements you got under the hypothesis of slope=0. If the P-value is small, less than 0.05, then the probability of the data set is very small and the hypothesis of zero slope is rejected. Such a small P-value is called statistical significance. This is taken as statistical proof that a non-zero rate of change has been observed. Statistical significance is indicative of a consistent rate of change. The value of the rate of change may be small, but its statistical significance may be used to prove the effectiveness of the BMP.

B.11 Number of Replicates

The formula for calculating the number of replicates is

$$N = 1 + \text{integer}\left[\frac{9.635 * s^2}{\Delta^2}\right]$$

where s=the standard deviation from the Baseline data, Δ is the difference between the mean from the Baseline data and the upper limit of the error range, and integer means to take the integer part of the results or to truncate the results. The effect of taking the integer part and adding one is to always round up.

For example, assume the Baseline data were the following:

35, 68, 49, 23, 84, 55, 47, 39, 61, 56.

Then,

n=10
$\bar{x} = 51.7$
s = 17.38.

If it is desired to know the mean to within ± 30%, say, then the upper limit is the mean plus 30% of the mean = 51.7 plus 0.3 * 51.7 = 67.21. So, Δ = 67.21 – 51.7 = 15.51. Inserting these values into the equation gives the required number of replicates as follows:

$$N = 1 + \text{integer}\left[\frac{9.635 * 17.38^2}{15.51^2}\right] = 1+\text{integer}(12.098) = 13.$$

Thus, 13 replicates are needed during the next year to determine the mean to within ± 30%.