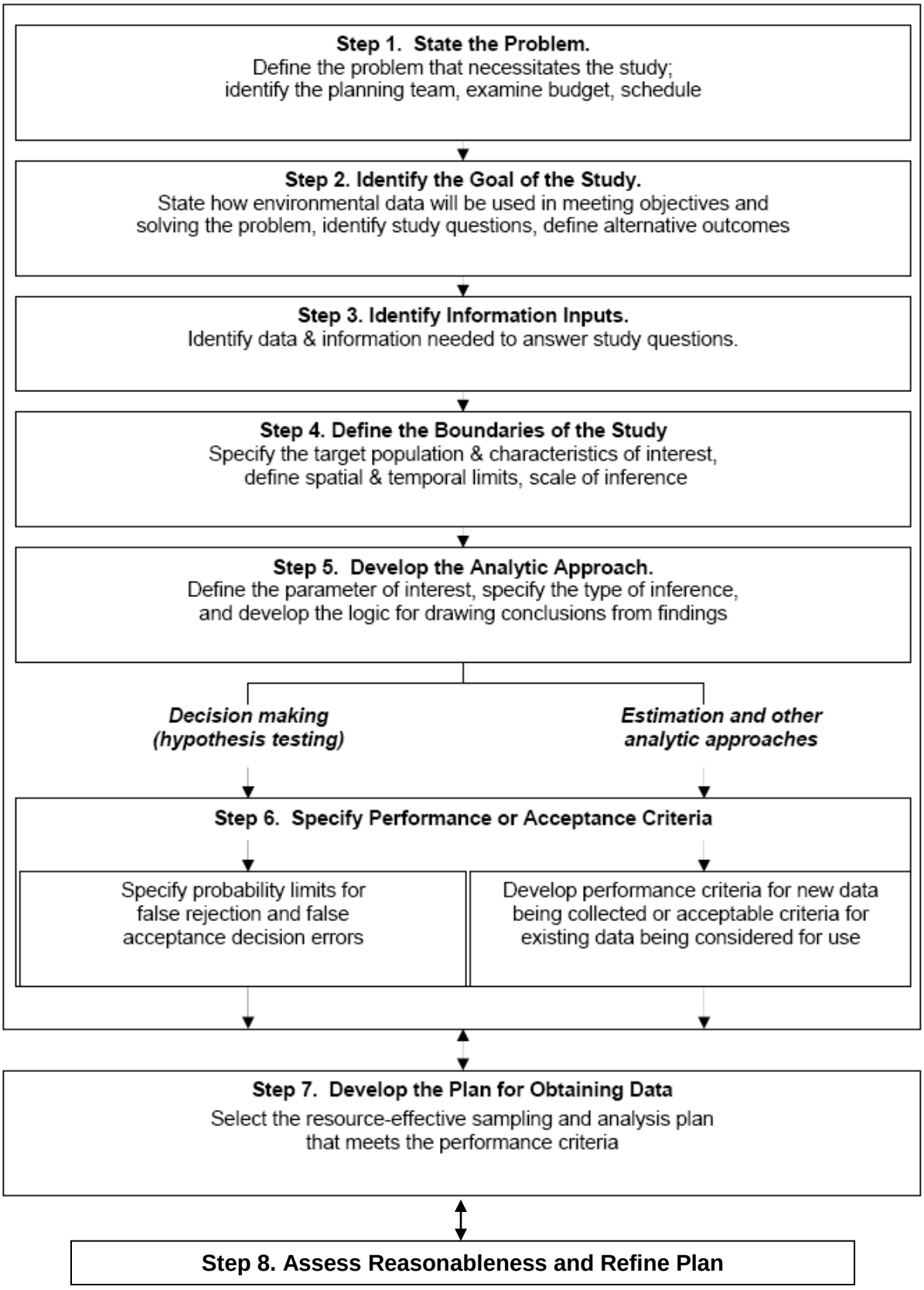
Chapter 2
DEVELOPING A MONITORING PLAN

This chapter provides a seven-step approach for developing a monitoring plan for collection of data to evaluate BMP effectiveness. The method described incorporates many elements that are included in EPA guidance regarding Data Quality Objectives (DQOs) (EPA 2006) and Quality Assurance Project Plans (QAPPs) (EPA 2002). Exhibit 2-1 provides a flow chart for working through these systematic steps, including the following:

1) Define Study Objectives
2) Identify Study Goals
3) Identify Information Inputs/Data Needs
4) Define Study Boundaries
5) Develop the Analytical Approach
6) Specify Performance or Acceptance Criteria
7) Develop Detailed Plan of Obtaining Data
8) Assess Reasonableness of Plan and Refine

This chapter describes representative considerations for each of these steps in the context of stormwater BMP monitoring.

Exhibit 2-1. Systematic Approach to BMP Monitoring Plan Development (EPA 2006)



2.1 Step 1. Define Study Objectives/State the Problem

It is very important that the objectives of a BMP monitoring program be clearly stated and recorded. The process of writing them down, working as a project team, generally results in careful consideration being given to the various possible objectives and questions to be answered. Written objectives help avoid misunderstandings by project participants, are an effective way of communicating with sponsors, and provide assurance that the monitoring program has been systematically planned.

Studies of BMP performance are usually conducted to obtain information regarding one or more of the following questions:

- What degree of pollution control or effluent quality does the BMP provide under normal conditions (i.e., representative storm types)?
- How does hydrology for developed conditions compare with pre-development hydrology in terms of peak flow rates, runoff volume, peak timing, site infiltration capacity, etc.?
- How does this performance vary from pollutant to pollutant?
- How does this normal performance vary with large or small storm events?
- How does this normal performance vary with rainfall intensity?
- How do BMP design variables affect performance?
- How does performance vary with different operational and/or maintenance approaches?
- Does performance improve, decay, or remain stable over time?
- Does performance vary seasonally? (For example, to what extent is infiltration reduced during cold temperatures?)
- How does this BMP's performance compare with the performance of other BMPs?
- Does this BMP help achieve compliance with water quality standards?

Many BMP monitoring programs have been established to satisfy requirements prescribed by permits to monitor the effectiveness of BMPs, but often the wording of such requirements is vague. Local program-specific objectives are likely to provide the soundest basis for planning a BMP monitoring study.

2.1.1 Key Activities for Step 1

Key activities for establishing monitoring objectives include the following:

1) Identify project team and identify decision makers, including project management/oversight, field staff, office staff, analytical technicians or laboratory, data users, project advisors, and peer reviewers.

2) Describe the problem and develop a conceptual model of the BMP to be investigated and identify general types of data that will be collected (i.e. hydrology, water quality, physical characteristics of facility). As part of this step, it is also important to understand and document basic site conditions affecting monitoring. For example, what is the source of the runoff (e.g., roofs, pavements) and where are the collection pipes? In general, what pollutants are anticipated from each source? For example, roof runoff may have copper from downspouts, pavement may have elevated chlorides from road salting in the winter, and lawns may have bacteria loading from geese or dogs. Are there unique soil conditions or karst topography?

   Exhibit 2-2 provides an example conceptual model of a bioretention cell and data that could potentially be collected including surface inflows; overflows/bypass; underdrain outflows; evapotranspiration; precipitation; infiltration to underlying water table; soil characteristics such as porosity, hydraulic conductivity, and pH; water quality constituent inflows, outflows, storage and bypass; and transport mechanisms including vegetative uptake, sorption/desorption, and advective transport. As Exhibit 2-2 illustrates, in many cases the conceptual model developed as a part of Step 1 will have many parameters, often more parameters than it will be feasible to measure.

Exhibit 2-2a. Conceptual Model for a Bioretention Cell with Underdrain
(Source: Hunt 2003)
(*diagram terms defined in Exhibit 2-2b*)

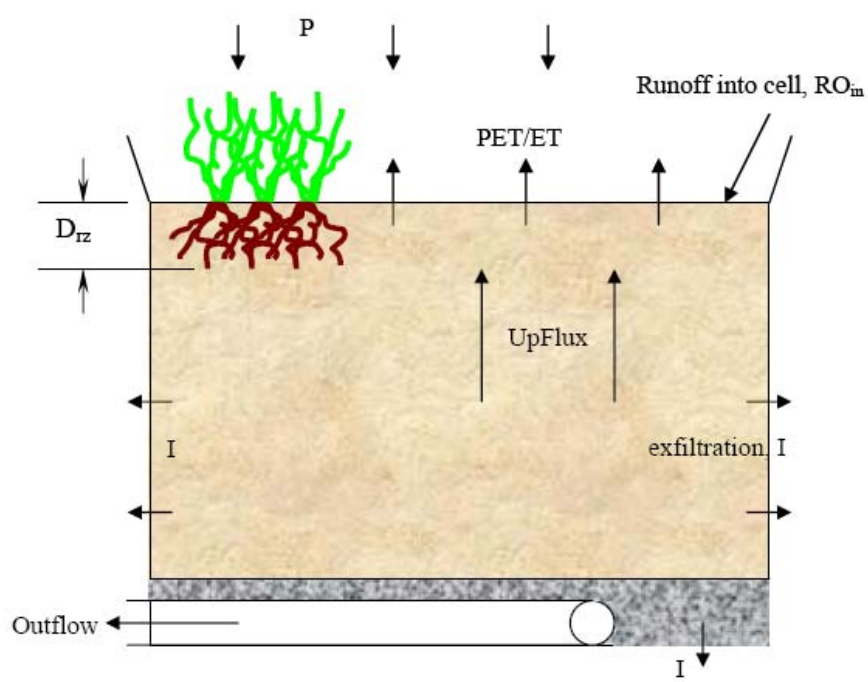


Exhibit 2-2b. Description of Conceptual Model Parameters (Source: Hunt 2003)

Parameter	Need/Purpose	Source of Data
Precipitation (P)	Input of water into the system. The precipitation record is adjusted to account for initial abstraction in this urban watershed.	Actual Record (for this model, the recorded values at the Greensboro rain gauge)
In-Situ Hydraulic Conductivity (K)	Exfiltration out the sidewalls and bottom of the bioretention cell account for a substantial portion of water loss in a bioretention device	Measured at site or County Soil Survey
Depth of Root Zone (r.z.)	During dry periods, this zone loses water until soils reach the wilting point. Initial rainfall must fill this zone before water passes through the media to the underdrains	DRAINMOD Reference Report (Skaggs, 1980)
Surface Water Content @ Saturation (WC_s)	Maximum amount of water in root zone	DRAINMOD Reference Report
Surface Water Content – lower limit (WC_{ll})	Minimum amount of water in root zone. This is set to be the wilting point of plants.	DRAINMOD Reference Report
Plant Coverage (in %)	Bioretention areas are not completely vegetated. This limits the amount of water stored in root zones.	Estimated at site.
Potential Evapotranspiration (PET)	An avenue through which water leaves the bioretention system. Precedent ET creates an initial sink for water in the bioretention device	Thornthwaite (1957). A worksheet within the model calculates PET for input into the model. This done on a monthly basis and is divided into hourly estimates during daylight hours.
Soil-Water Characteristic	Relates suction pressure to amount of water retained in the soil media.	Lab samples measured in laboratory with pressure plates.
Drainage Volume-Water Table Depth Relationship	Relates water table rise and fall to volume of water leaving the soil profile	Calculate from Soil-Water Characteristic Curve or Standard Table
Upward Flux	Determines if ET is limited by depth of water	WTD- Upward Flux relationships in literature. DRAINMOD Reference Report used here.

3) When a complex conceptual plan with many parameters is developed, it is necessary to make an assessment of the level of complexity of the monitoring that will be conducted. As a part of Step 1, parameters in the conceptual model can be prioritized into tiers that will aid in further steps to determine which parameters will be monitored, which ones will not be, and which ones will be estimated or calculated.

4) Discuss and evaluate alternative approaches to evaluating problem. Can goals be accomplished by evaluating data that have already been collected for other projects? If additional data are collected, can they be complemented by already-collected data? Because of the typically high cost of BMP monitoring, it may be desirable to evaluate alternative means for addressing some information needs (assuming that BMP monitoring

is not required to comply with a permit). Depending on the situation, sediment sampling, biological sampling, and/or visual surveys of the stormwater conveyance system may be cost-effective alternatives to stormwater quality monitoring. Literature reviews may also help address some stormwater management issues.

5) Identify key resources, constraints, deadlines. Resources could include the number of candidate BMP sites for monitoring, availability and/or cost of monitoring equipment, availability of staff to assign to the project, technical expertise of staff, funding resources, partnership opportunities, etc. Constraints typically arise from physical factors, practical factors and lack of resources. Constraints can include practical ability to collect measurements/samples; achievable accuracy and precision of measurements; geographic proximity of monitoring site to laboratory and/or location of monitoring staff; wet-weather site access; sample collection and processing time; natural variability in stormwater runoff hydrology and water quality; financial limitations; and other factors such as the age of the facility (e.g., are conditions at a new facility adequately stabilized to represent expected normal site conditions?). Deadlines may arise for many reasons including permit compliance timelines, grant durations, seasonality or internal progress goals.

Typical information that should be collected as a part of assessing resources in Step 1 includes the following:

- Results from prior surface water and groundwater quality studies, other BMP monitoring studies in the local area, sediment quality studies, aquatic ecology surveys, dry weather reconnaissance, etc.
- Drainage system maps.
- Land use maps (or general plan or zoning maps).
- Aerial photographs.
- Precipitation and streamflow records.
- Reported spills and leaks.
- Interviews with public works staff.
- Literature on design of structural BMPs to understand functionality and pollutant removal processes.

For BMPs monitored in industrial areas, the following information may also be relevant:

- BMP performance data for similar industries in region.
- Facility map(s) showing locations of key activities or materials that could be exposed to stormwater.

- Lists of materials likely to be exposed to stormwater.
- Reported spills and leaks.
- Interviews with facility staff and others who are knowledgeable about the facility.

2.1.2 Outputs from Step 1

Primary outputs from this Step 1 include the following:

1) A concise written description of the problem to be investigated. This can later be incorporated into the QAPP directly.

2) A conceptual model of BMP operation and site conditions including key parameters for evaluation of BMP effectiveness. Representative practical considerations (i.e., "reality checks") associated with this step that can influence the financial and scientific success of the project are summarized in Exhibit 2-3.

3) A roster and conceptual organization chart for the project team.

4) A summary of resources including estimated available budget, available staff, potential partnerships, key deadlines and preliminary schedule.

Exhibit 2-3. Representative Practical Considerations at the Project Planning Level (Source: Granato 2009)

Basic Meteorological Data:

- Expected number of storms/year—will it be possible to get an adequate sample set during the time frame of the study?
- Expected range of sampling rates based on anticipated range of volumes per storm—will it be possible to get enough flow at the inlet to measure the flow and enough sample volume for multiple analyses without constant adjustments in sampling rates?
- What range of storm volumes associated with which storms should be sampled?
- Are average storm durations and holding times for proposed analytes realistic? (For example, if transporting bacteria samples to a lab within a 4-6 hour holding time is an objective and the average storm duration is 18 hours, do you have the time and people to do this sampling?)

Basic Watershed Characterization ("reality check" parameters):

- Imperviousness of tributary area—what proportion of precipitation is realistically expected as runoff at the inlet?
- Land use—what range of concentrations and constituents are expected? For example, it may be helpful to develop a basic range of expected water quality results based on sources such as the National Stormwater Quality Database (http://unix.eng.ua.edu/~rpitt/Research/ms4/Paper/Mainms4paper.html) for estimates.

Basic Engineering Design Information

- Is the selected site representative of good design practice or is the selected BMP an "outlier"?
- What is the expected range of storm sizes that will cause bypass, overflow, or failure? If bypass or overflow is anticipated, has monitoring equipment been planned for those locations?
- What design aspects will affect monitoring design? For example, how long will the outflow discharge need to be monitored to obtain an outflow sample comparable to the runoff inflow?

2.2 Step 2. Identify Study Goals

The primary focus of Step 2 is to further define the overall study objective and problem statement into a series of more-detailed questions that can then be evaluated for decision making implications and evaluated to determine key measurements that must be collected. For simple evaluations, a study goal may be summarized in a single question. For more complex studies,

typical of many BMP monitoring situations, multiple questions will be required to define study goals. Step 2 should further define the focus of Step 1 and specify additional details on data collection plans and the specificity of the question to be answered by the data. Examples of goals/questions formulated in Step 2 based on objectives formulated in Step 1 could include the following:

Example Step 1 Objective:

1) What degree of pollution control or effluent quality does a bioretention BMP provide under normal conditions?

Example Step 2 Goals/Questions:

1) For all storm events monitored, do water quality sampling results from inflows and outflows exhibit a normal distribution? Do they exhibit a log-normal distribution?

2) On an annual basis, what fraction of runoff bypasses the bioretention BMP?

3) What fraction of the annual load of total phosphorus bypasses the BMP on an annual basis?

4) For all storm events monitored, what are the measured differences in BMP surface inflows and outflows in terms of peak rate and runoff volume?

5) For storm events up to the BMP design storm, what are the measured differences in BMP surface inflows and outflows in terms of peak rate and runoff volume?

6) For total rainfall depths up to the 90th percentile storm, are inflow and outflow concentrations of total phosphorus statistically significantly different at a 95 percent confidence level?

7) How does the median underdrain outflow total phosphorus concentration from the bioretention BMP compare with the median runoff value from the National Stormwater Quality Database?

8) How does the median underdrain outflow total phosphorus concentration from the bioretention BMP compare with the median values in the inflow?

These are but a few of the many goals that could be specified to evaluate the degree of pollution control or effluent quality of a bioretention BMP under normal conditions. The number of questions that can be evaluated and project goals are typically limited by cost, staff, schedule, and the number of rainfall events that can reasonably be expected during the study period.

2.2.1 Key Activities for Step 2

Key activities for identifying and defining study goals include the following:

1) Identify principal study question or questions and evaluate responsive actions to various outcomes resulting from answering study questions. For some BMP effectiveness monitoring projects, there may not be a near-term direct response that is directly related to the data collected. Data may be submitted to the BMP Database and ultimately may contribute to decisions regarding BMP design criteria. On the other hand, some BMP effectiveness monitoring projects may lead to responsive actions based on results of monitoring, especially when permit compliance is concerned.

   For example, if a goal of a study is to assess whether a constructed wetland basin has effluent quality that is at or below the receiving water standards for dissolved copper, and results indicate that the pond effluent concentrations are higher than the standards, responses could include modifications to the pond outlet to increase residence time, consideration of alternate supplemental treatment, collection of additional data to tighten confidence intervals, etc.

2) Develop a decision statement or estimation statement based on potential outcomes or answering principal study question(s). EPA DQO guidance defines two primary types of problems that a monitoring study may seek to address: (1) decision problems and (2) estimation problems (EPA 2006). Decision problems generally are framed so that the results of the monitoring will result in an action, whereas estimation problems are often framed so that the results of the monitoring will provide additional information to more accurately describe the system and/or conceptual model. Examples of decision and estimation problems and statements for BMP effectiveness monitoring are provided in Examples A and B, respectively.

   <u>Example A. Decision Problem/Principal Study Question</u>

   - Does the mean annual nitrate concentration from an extended detention basin-constructed wetland treatment train exceed the agreed-upon concentration from the development agreement with the downstream homeowner's association for protection of lake water quality?

   <u>Example A. Decision Statement/Alternative Actions</u>

   - Concur that mean annual nitrate concentrations from the extended detention basin-constructed wetland treatment train meet the agreed upon concentration from the development agreement with the downstream homeowner's association for protection of lake water quality.

   - Analyze inflow nitrate data to look for unusually high values that could indicate source.

   - Survey homeowners on fertilization practices. Restrict fertilizer application.

   - Reevaluate maintenance schedule for constructed wetland to determine if BMP seasonally exports nitrate.

- Consider design modifications to pond or wetland basin.

Example B. Estimation Problem/Principal Study Question:

- Does the effluent quantity and quality for turbidity from a green roof improve as vegetation becomes better established?

Example B. Estimation Statement/Alternate Actions:

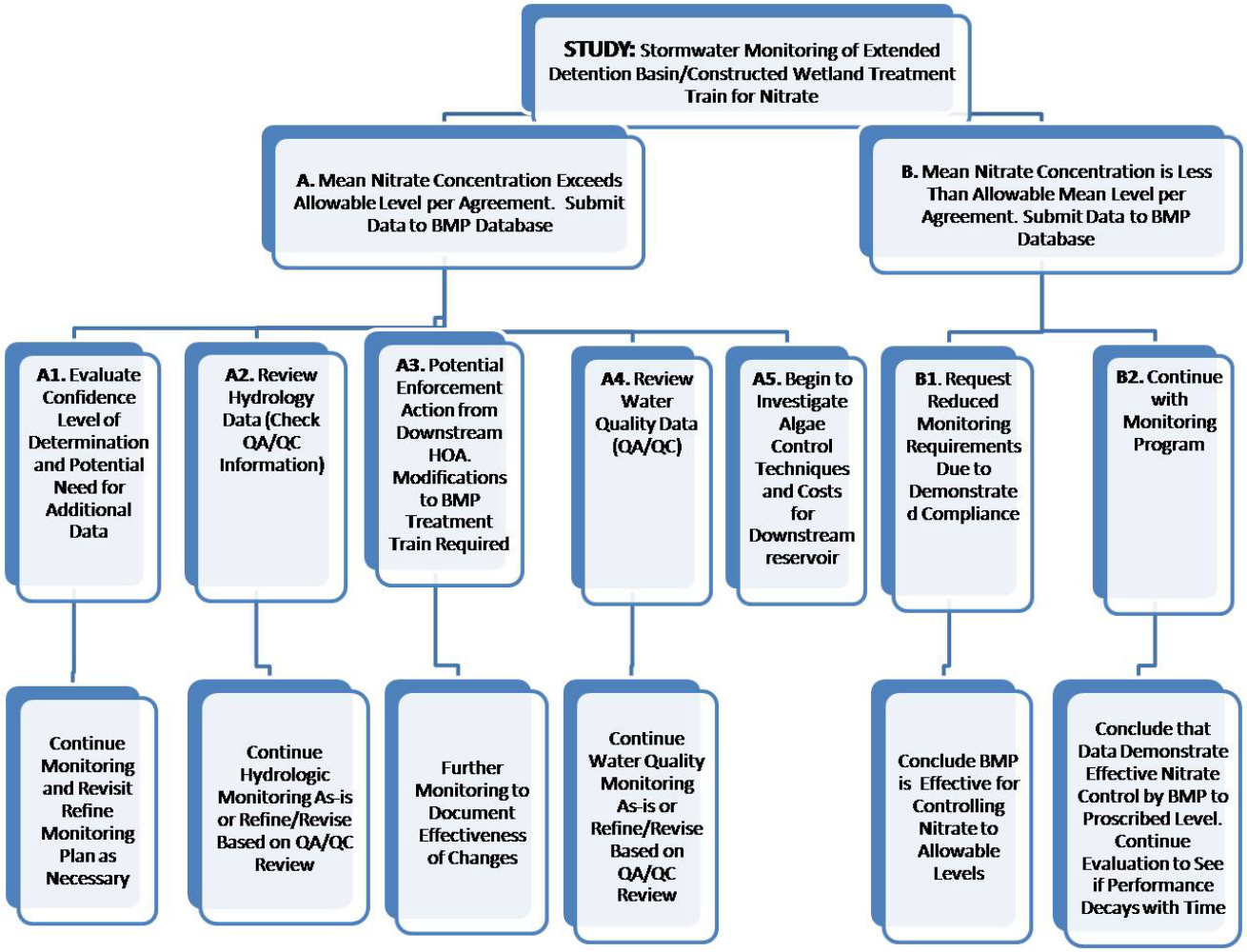
- Based on study from initial planting through three years of growth, runoff volume and turbidity levels both decline, indicating performance improves over time. Consider further evaluation of seasonal trends, effects of different types of vegetation, different types of soils, etc.

- Data from three year investigation show decrease in runoff volume but increase in turbidity levels. Evaluate hydrology/precipitation for trends, quantify extent of vegetative growth/exposed soil, consider further sampling/alternate sampling design to isolate source of turbidity.

- Study indicates that over three year study period, runoff volume and turbidity yields from the green roof increase. Evaluate hydrology/precipitation to see if increase runoff volume is due to greater precipitation, evaluate potential sources of turbidity.

Prioritize decision and/or estimation outcomes based on their relation to each other and relation to overall study goal. This task provides up-front planning regarding "what to do with the study results" under different scenarios. Using the decision-based example above, Exhibit 2-4 illustrates prioritization of outcomes. As shown in this example, two different series of actions result depending on the study results.

Exhibit 2-4. Example Prioritization of Outcomes and Response



2.2.2 Outputs from Step 2

Primary outputs from Step 2 include the following:

1) Well defined principal study questions. Questions should provide a refinement of the general study objective defined in Step 1 and provide detailed information that can be used in further steps to identify specific data to collect.

2) List of alternative outcomes resulting from answering study question. Anticipate conclusions that may be drawn from data collected and data analysis techniques employed.

3) Decision statement or estimation statement describing how findings of monitoring will be used to aid in decision-making and fulfill objective from Step 1.

2.3 Step 3. Identify Information Inputs

Step 3 identifies the specific type of information and data needed to address the overall study objective identified in Step 1 and the more detailed goals that have been developed as a part of Step 2. This step combines the conceptual model developed as a part of Step 1 with the specific questions of Step 2 to determine which components of the conceptual model will be measured, which will be estimated, and which may be available from existing technical literature. Constraints including available funding, equipment, technical expertise, practical considerations and capabilities for calculating or estimating various parameters from the conceptual model will shape the methods used to obtain the information inputs that the monitoring study seeks.

2.3.1 Key Activities for Step 3

Activities to identify information inputs (Step 3) include the following:

1) Determine types and potential sources of information needed. Potential sources of information can range from previously conducted studies to new data collection. Due to the high costs of data collection, garnering as much information as possible from past monitoring efforts located in the region or focused on similar BMPs can result in significant costs savings. Review of literature can provide comparable data, reveal problems encountered in similar monitoring efforts and troubleshooting solutions, and define criteria and measurement methods so that additional data collected will be comparable with past data collected.

   For data that are being collected that ultimately will be entered into the BMP Database, the data entry sheets from the BMP Database (http://www.bmpdatabase.org) can be a valuable resource for determining which facility characteristics, hydrologic and water quality parameters should be measured or estimated. Appendix A provides a copy of these forms.

2) Determine the basis for specifying performance or acceptance criteria for the collected data. Performance and acceptance criteria for data will vary with project objectives and goals and with the type of data collected (i.e., estimates from literature, reported values for measurements collected by others in relatively close proximity, direct measurements. Considerations may include the following:

   - Literature values: Considerations may include regional proximity to the study site, similarities in annual precipitation, similarities in BMP design criteria, similarities in land use, similarities in soil characteristics in the watershed and underlying the BMP, monitoring methods and quality assurance/quality control (QA/QC) protocols, ability to contact and communicate with the researchers who collected the data, and other factors.

   - Measurements collected by others (e.g., precipitation data collected at nearby weather station, stream gaging information, etc.): considerations may include the agency/entity collecting the data, proximity to the study site, precision and accuracy of the data, QA/QC protocols, temporal and spatial resolution of data, and similarities of environment in which data are collected and study environment.

   - Direct measurements: Considerations include precision and accuracy of monitoring equipment, method detection limits for analytical methods, sampling collection methods, allowable holding times, requirements for duplicate sampling and analysis and other QC checks (spikes, blanks, etc.).

   In addition to criteria for specific measurements, this step should also identify criteria for storm events that will be sampled. (See Section 2.3.2 for a more detailed discussion of this important topic.) These criteria include the number of storms to be monitored during the study; storm characteristics such as minimum or maximum precipitation depth, storm intensity, duration for inclusion in the study; antecedent dry time; and/or relationship of monitored storms to "typical" event (i.e., collect samples from storms with an event precipitation depth that is plus or minus one standard deviation from the mean annual storm depth for the area). In most cases, monitoring programs will focus on commonly occurring storms, typically serving as the basis of BMP design rather than large, in frequent storms. Although time and cost constraints may limit the ability to monitor infrequently occurring storm, data collected for infrequently occurring storms are often valuable for characterizing the limitations of BMPs. Even in large events, BMPs may provide treatment for the initial runoff from the event, providing water quality benefits. It also may be useful to document the dilution that can occur in larger events (for example, the quality of bypassed runoff in a larger event may be better than the treated portion, especially if remobilization of previously captured pollutants occurs within the BMP).

3) Verify the availability of appropriate sampling equipment and analysis methods. The project team should develop and review the list of desired measurements and specify how each measurement will be collected. This will typically involve

researching a variety of stormwater sampling products ranging from grab sampling equipment to automated sample collection equipment. For infiltration oriented practices, be aware that groundwater and/or vadose zone monitoring may be important components of the monitoring program, depending on study objectives. For any electronic equipment used, two very important considerations are the power source and communications capabilities (i.e., modem). This step will also involve contacting local laboratories to determine analytical methods that they may use for sample analysis, detection limits available for study parameters, and lab certification status. For constituents such as TSS and filtered metals that are subject to significant variability, it is important to thoroughly review and document lab procedures, even for certified labs following standard methods.

2.3.2 Determining Number of Storms Needed for Meaningful Statistical Assessment

2.3.2.1 General Considerations

The number of storms to be monitored each year (i.e., monitoring frequency) and the representativeness of the monitored storms with regard to study objectives are important considerations in planning the monitoring program. For example, a common LID study objective may be to evaluate performance of LID technique for small, frequently occurring storms, as opposed to large storms that exceed design parameters.

One of the most frequently overlooked factors in designing a monitoring plan is the number of samples required to obtain a statistically valid assessment of water quality. Budget and staff constraints generally limit the number of storms, locations, and parameters to be monitored. Program objectives should be weighed in light of available resources to determine the best mix of monitoring frequency, locations, and parameters. The cost of learning more (i.e., conducting more intensive monitoring) should be compared to the cost implications of moving forward too quickly and implementing extensive controls before having learned enough to guide planning, stormwater management commitments, and/or negotiations with regulatory agencies. The cost of controlling unimportant pollutants and/or unimportant sources, or implementing ineffective BMPs could easily exceed the cost of monitoring to learn more about actual BMP performance under the conditions that prevail in the system. Clearly, there is a need for balance here, because endless studies should not be substituted for control actions. In general, however, many measurements (i.e., many samples during many events) are necessary to obtain enough data to be confident of actual BMP performance, as opposed to drawing erroneous conclusions based on “noisy data” (e.g., variability artifacts caused by external factors, equipment and operator errors). Consequently, BMP effectiveness studies can be expensive and time-consuming.

Stormwater quality may vary dramatically from storm to storm. Therefore, monitoring a large number of storms is required if the objective of the program is to obtain accurate estimates of stormwater pollution in a given catchment (e.g., to determine whether water quality is changing over time or whether a given BMP is effective). However, staff and budget constraints typically limit monitoring to either a smaller set of parameters for

many storms, or a more detailed monitoring approach including a larger set of parameters for a few storms. Often goals for a monitoring effort (e.g., to demonstrate that a specific BMP is achieving a given level of removal of a constituent) may not be consistent with fiscal limitations of the project.

Four factors influence the probability of identifying a significant temporal and/or spatial change in water quality:

1) Overall variability in the water quality data.

2) Minimum detectable change in water quality (difference in mean concentration).

3) Number of samples collected.

4) Characteristics of storms sampled (e.g., do the samples focus on frequently occurring events, or a wide range of storm conditions).

5) Desired confidence level from which to draw conclusions.

Estimates of the number of samples required to yield statistically valid monitoring results are necessary for making decisions about the nature and extent of monitoring efforts prior to implementation. Statistical analysis may be conducted to estimate how many events need to be monitored to achieve a desired confidence in a conclusion (i.e., power analysis). Performing a power analysis requires that the magnitude of detectable change, the confidence level, and the statistical power or probability of detecting a difference are defined. Typically, the confidence level and power are at least 95 percent and 80 percent, respectively, meaning that there is a 5 percent probability of drawing an incorrect conclusion from the analysis and a 20 percent probability that a significant change will be overlooked.

2.3.2.2 Sampling Strategies and Determination of Number of Samples for Paired Evaluations[1]

The comparison of paired data sets is commonly used when evaluating the differences between two situations (locations, times, practices, etc.). One common situation is the collection of paired influent and effluent samples from a control device being evaluated. However, it is critical to ensure that the samples being collected are truly "pairs." The following discussion presents a range of sampling situations where paired analyses may be desired.

When monitoring very "small" control devices (those that have a short hydraulic detention time within the device), it is possible to take paired samples on a real-time basis. This would be possible for a flow-thru screen, for example, where there is no storage in the device. In this example, instantaneous pairs of influent and effluent samples can be taken because the same water is represented in both samples.

[1] Guidance in this section taken directly from "Effort for Paired Samples." Unpublished text prepared by Robert Pitt, P.E., Ph.D., University of Alabama, March 2009.

For somewhat "larger" devices (such as catchbasins with sumps, or an inlet insert placed in a catchbasin sump), the hydraulic residence time can be calculated using the discharge rate and the storage volume, as shown below:

Sump storage volume: 52 cubic feet
Discharge rate: 120 gal/min (= 16 ft^3/min = 0.27 ft^3/sec)

Residence time = 52 ft^3/16 ft^3/min = 3.3 minutes

Therefore, the flow would have to be reasonably steady for at a time substantially longer than this residence period, as this would likely be a completely mixed system and a plug flow system. Therefore, several residence time periods would be needed before approximate steady state conditions are attained. During this longer period, the influent may be rapidly changing (such as during summer thunderstorms), while the effluent will be slightly moderated by the storage in the system. A paired sampling scheme under this scenario could be comprised of many subsamples taken over a moderate period of time (several detention periods), such as 10 to 30 minutes, and composited for single influent and effluent analyses. Therefore, a single inlet and outlet sample would be a single pair, and each would be comprised of many composited subsamples taken over several residence time periods. In this case, an example could involve manual dipper samples every 30 to 60 seconds for a 30 minute sampling period.

A similar scenario would include a grass swale where the travel time along the test reach may be several minutes. However, in this case, plug flow would be more likely than completely mixed flow. Paired samples can also be composited over several travel time increments between the sampling locations, but it would also be a good idea to delay the start of the effluent samples by the expected travel time.

For even larger devices where the residence time is longer, such as in a bioretention device where the water may be in the device for many minutes to a few hours (assuming a small amount of ponding followed by filtering through soil media and then collected in an underdrain), composite samples should be taken during the complete period of flow at the influent and effluent (underdrain and overflow bypass) for each event. Therefore, each event would result in a single pair of samples.

For stormwater control devices that may contain water for extended periods, such as a wet detention pond or a wetland, the effluent during a rain event may be quite un-related to the influent water. During small rains, the total event runoff volume may be much smaller than the storage volume of the device. In this case, the effluent may be mostly "historical" water that has resided in the control device for some time. For very large rains, the runoff volume may be several times the storage volume of the device, and the effect of the "dilution" of the stored resident water would be less problematic. Therefore, paired sampling for these devices is much more complicated. Single event pairs are possible when the runoff volume is large compared to the storage volume. It is a good idea to calculate this "flushing ratio" for the events before analyzing the data. In this case, the best approach may be to examine the sum-of-loads for a season, where most of the influent and effluent during an extended period have been represented. Within this

season, periodic large events can be evaluated as pairs, if the flushing ratio is large. It is important that all sizes of events be included in a sampling program, as small events are of interest when evaluating permit limits for heavy metals and bacteria, for example. Performance as a function of storm event size (flushing ratio) is also important and can be directly determined.

The number of influent/effluent sample pairs needed can be calculated based on the data quality objectives of the study, in a similar way that experimental design sampling efforts for characterization can be determined. Once the confidence and power of the desired outcome is known, it is possible to determine the sampling effort to identify a specific difference in influent and effluent concentrations. If only large differences need to be identified, then fewer samples are needed. However, if lower levels of performance need to be quantified, then more samples are needed. This "detection limit" of performance needs to be identified for the experimental design effort. Most statistical analyses determine if the influent and effluent are significantly "different." However, many misinterpret the results if the samples are not statistically different by saying that they are the "same." The correct interpretation is that there were not enough samples taken to show that they were significantly different. It is theoretically possible to statistically show that any paired set of samples are different if enough samples are taken. Of course, economics and opportunities limit the sampling effort. However, it is possible to identify the expected limit that any experimental design can detect as part of the sampling.

An equation (similar to the one used for single point characterization) that can be used to estimate the needed numbers of samples for a paired comparison (Cameron, undated; Burton and Pitt 2001) is as follows:

$$n = 2\,[(Z_{1-\alpha} + Z_{1-\beta})/(\mu_1 - \mu_2)]^2\sigma^2$$

where:

n = number of sample pairs needed

α = false positive rate ($1-\alpha$ is the degree of confidence. A value of α of 0.05 is usually considered statistically significant, corresponding to a $1-\alpha$ degree of confidence of 0.95, or 95%)

β = false negative rate ($1-\beta$ is the power. If used, a value of β of 0.2 is common, but it is frequently ignored, corresponding to a β of 0.5.)

$Z_{1-\alpha}$ = Z score (associated with area under normal curve) corresponding to $1-\alpha$

$Z_{1-\beta}$ = Z score corresponding to $1-\beta$ value

μ_1 = mean of data set one

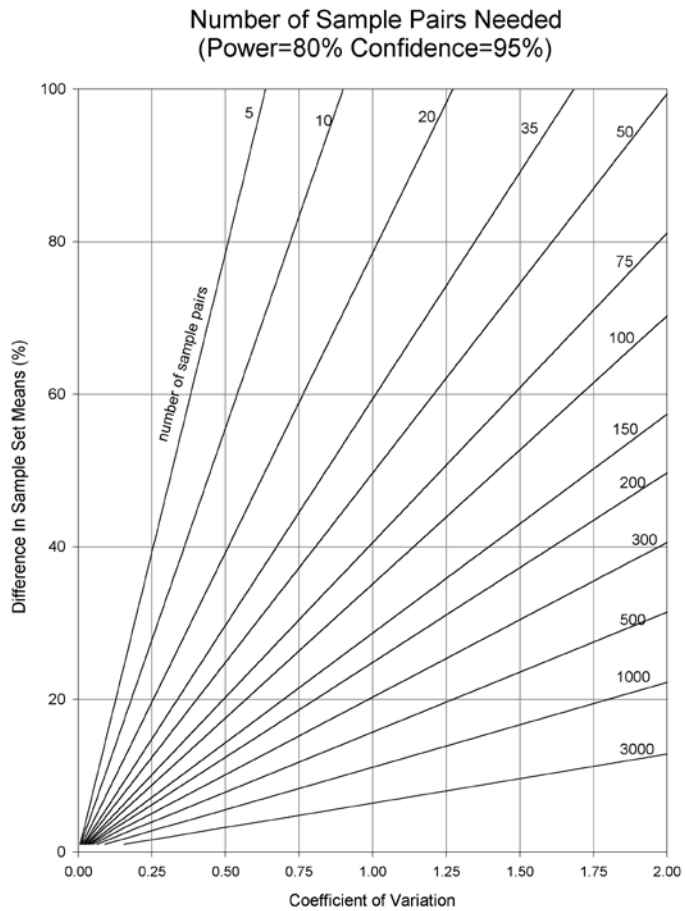
μ_2 = mean of data set two

σ = standard deviation (same for both data sets, same units as μ. Both data sets are also assumed to be normally distributed.)

This equation assumes that the two data sets be normally distributed and have the same standard deviations. Most stormwater parameters of interest are likely closer to being log-normally distributed. If the coefficient of variation (COV) values are low (less than about 0.4), then there is probably no real difference in the predicted sampling effort. This method should be used using log-transformed data for the more likely expected higher COV conditions.

Exhibit 2-5 (Pitt and Parmer 1995) is a plot of this equation (normalized using COV and differences of sample means) showing the approximate number of sample pairs needed for a typical value of α of 0.05 (degree of confidence of 95%), and a β of 0.2 (power of 80%). As an example, twelve sample pairs will be sufficient to detect significant differences (with at least a 50% difference in the parameter value) for two locations, if the coefficient of variations are no more than about 0.5. Appendix D (Pitt and Parmer 1995 and Burton and Pitt 2001) contains similar plots for many combinations of other levels of power, confidence and expected differences. References such as the BMP Database (http://www.bmpdatabase.org) and the National Stormwater Quality Database (http://rpitt.eng.ua.edu/Research/ms4/Paper/Mainms4paper.html) can be referenced for representative COVs when using Exhibit 2-5.

Exhibit 2-5. Sample Effort Needed for Paired Testing at a Power of 80% and Confidence of 95% (Pitt and Parmer 1995; Burton and Pitt 2001)



2.3.3 Outputs from Step 3

Primary outputs from the identification of information needs process (Step 3) include:

1) List of environmental characteristics, measurements and estimates that will provide answers to questions formulated for study goals from Step 2 and overall study objective from Step 1. For example, in a side-by-side small watershed comparison where one site has extensive rain gardens and the other does not, environmental characteristics, measurements and estimates could include the following:

   - Watershed area (each watershed).
   - Watershed imperviousness and land uses (each watershed).
   - Area of rain gardens in watershed.

- Design characteristics of rain gardens (hydrologic and hydraulic sizing, soils, vegetation, underdrains/no underdrains, etc.).

- Stormwater runoff hydrographs at primary storm sewer outfalls from each watershed at 5-minute intervals.

- Precipitation data (5-minute intervals) from tipping bucket rain gages in the watersheds.

- Flow-weighted composite water quality samples from each outfall for all events producing measurable runoff that can be sampled (acknowledging that there will be some events where staff errors or equipment malfunction will result in missed samples).

- Chemical analysis by qualified laboratory for total phosphorus, total suspended solids and chemical oxygen demand. (The laboratory selected could be a certified commercial laboratory or a well qualified research laboratory reporting complete QA/QC information.)

2) Information on the number of variables that will need to be evaluated and measurements collected. Depending on the objectives and goals of the study and the statistical confidence desired (as discussed in Section 2.3.2), the project team can identify the number of measurements desired. Due to both human and equipment errors and the extreme variability of stormwater runoff events, it is good practice to plan on collecting samples from 10 to 20 percent more events than are indicated from statistical analysis of data needs. By planning for these additional events, budget problems can be minimized due to "missed" events and progress expectations will be more reasonable.

3) Type and quality of information that will be needed to meet data performance or acceptance criteria. This will include a listing of the required measurement accuracy, method detection limits for analytical methods, and temporal and spatial resolution of data. It is noteworthy that detection limits achievable for research purposes may be lower than those available in commercial laboratories. In addition, available detection limits may vary from laboratory to laboratory (and even within a given lab, depending on cost). Therefore, selection of analytical methods and acceptable detection limits should take into consideration receiving water quality standards and other study objectives.

4) Identification of appropriate sampling and analysis methods. For a stormwater BMP effectiveness monitoring program, there are many alternatives for sampling and analysis that should be specified in this step including the following:

- Manual sampling versus automatic sampling.

- Grab sampling versus composite sampling (flow-weighted, time-proportional, random).

- Precipitation measurement (tipping bucket gage, total precipitation gage, nearby weather station).

- Flow measurement methods and equipment.

- Types of sample containers (and collection tubing) for various parameters.

- Definition of storm events (e.g., 6-hour separation or required antecedent dry condition). It may be useful to specify an antecedent dry period (i.e., 72 hours), as well as preceding/proceeding rainfall amounts. For example, it may not be useful to analyze samples from a 0.10-inch event occurring three days after a 0.5-inch event, but it could be useful to analyze samples from a 0.5-inch event within 24 hours of a 0.1-inch event. (Note: Some researchers prefer to sample all events so that a more complete population of sample conditions is represented; however, to accommodate a greater number of analyzed storms, it may be necessary to make economic tradeoffs with the number of parameters tested for some storms.)

- Allowable holding times.

- Handling requirements including field filtration, sample preservation, and temperature control.

- Field measurements such as pH, dissolved oxygen, oxidation-reduction potential, temperature, conductivity.

- Laboratory analyses and detection limits. Often there will be multiple methods for a given parameter with different detection limits and potentially different sample handling and preservation requirements. The method selected should have a detection limit that is consistent with the study goals and objective.

2.4 Step 4. Define the Boundaries of the Study

2.4.1 Key Activities for Step 4

Key activities for Step 4 include the following:

1) Define the target population of interest and relevant spatial boundaries. For stormwater BMP effectiveness monitoring, the target populations will generally include site/facility characteristics, hydrologic parameters and water quality parameters. The spatial boundaries of BMP monitoring efforts are critically important, especially for monitoring efforts where a mass balance estimate is desired for the BMP. Practical considerations in defining spatial boundaries of a monitoring study or evaluating a site to determine if it exhibits spatial boundaries that are likely to yield useful data include:

- The storm drain system should be sufficiently well understood to allow a reliable delineation and description of the catchment area (e.g., geographic extent, topography, land uses).

- For monitoring stations that will be used to measure flow in open channels, the flow measurement facilities need to be located where there is suitable hydraulic control so that reliable rating curves (i.e., stage-discharge relationships) can be developed. In other words, the upstream and downstream conditions must meet the assumptions on which the measurement method is based.

- Where possible, stations should be located in reaches of a conveyance where flows tend to be relatively "stable" and "uniform" for some distance upstream (approximately 6 channel widths or 12 pipe diameters), to better approach "uniform" flow conditions. Thus, avoid steep slopes, pipe diameter changes, junctions, and areas of irregular channel shape due to breaks, repairs, roots, debris, etc.

- Achieving well mixed conditions at a sampling station is an important consideration. In some cases, it may be advantageous to construct a “sampling box” to create cascading flow to the sample intake, thereby improving mixing.

- Locations likely to be affected by backwater and tidal conditions should be avoided since these factors can complicate the reliable measurement of flow and the interpretation of data.

- Stations in pipes, culverts, or tunnels should be located to avoid surcharging (pressure flow) over the normal range of precipitation.

- Stations should be located sufficiently downstream from inflows to the drainage system to better achieve well-mixed conditions across the channel and to favor the likelihood of "uniform" flow conditions.

- Stations should be located where field personnel can be as safe as possible (i.e., where surface visibility is good and traffic hazards are minimal, and where monitoring personnel are unlikely to be exposed to explosive or toxic atmospheres).

- Stations should be located where access and security are good, and vandalism of sampling equipment is unlikely. In some cases, it will be necessary to secure sampling equipment in protective boxes to guard against vandalism.

- Stations should be located where the channel or storm drain is soundly constructed. If a BMP is being installed and monitored for regulatory purposes, incorporating sampling needs into the BMP design can be cost-effective, as opposed to trying to retrofit the sampling equipment to the BMP later.

- If an automated sampler with a peristaltic pump is to be used, and the access point is a manhole, the water surface elevation should not be excessively deep (i.e., it should be less than 6 meters, or 20 feet, below the elevation of the pump in the sampler, and preferably less than 4.5 meters or 15 feet deep).

- If automated equipment is to be used, the site configuration should be such that confined space entry (for equipment installation, routine servicing, and operation) can be performed safely and in compliance with applicable regulations.

Prior to making a final determination on sample locations, each potential sampling station should be visited, preferably during or after a storm to observe the discharge. A wet-weather visit can provide valuable information regarding logistical constraints that may not be readily apparent during dry weather. Photo documentation of these conditions can be very helpful during implementation of the monitoring program. The importance of knowing the site cannot be overstated. This also includes being aware of factors influencing the site such as french drains, illicit discharges and maintenance practices such as street sweeping and power washing.

2) Define what constitutes a sampling unit. A sampling unit typically consists of a specified volume of an environmental media (water or soil) collected for analysis. The volume will be dictated by the volume requirements of the analytical methods used. For BMP effectiveness monitoring programs, sampling units could range from a series of discrete samples, to a flow-weighted composite sample. When considering what constitutes a sampling unit, it may also be useful to specify what constitutes a "complete" sampling event for a site. It is common to have multiple inflows to a BMP, and situations where some sampling stations function properly for sample collection and others malfunction should be anticipated. For example, to have a "complete" sampling event, is it necessary that all sampling stations at the site collect samples? Alternatively, would it be acceptable to count an event as "complete" if two of three inflow samplers trigger and collect samples along with the outflow? Answering these types of questions upfront is important because the cost of laboratory analysis can be significant and analysis of samples from storm events that do not meet criteria for being "completely" monitored may not be worthwhile to analyze in some cases.

3) Specify temporal boundaries and practical constraints associated with sample/data collection. Temporal boundaries include the start date of the BMP monitoring effort, the duration of the monitoring program, antecedent dry time required before samples are collected (typically 72 hours—but also dependent on the preceding amount of rainfall and intensity/washoff potential), frequency of sample or sub-sample collection during an event, allowable sample holding times, and targeted storm durations for monitoring. Temporal constraints may include access restrictions to a monitoring location during non-daylight hours, operating hours of laboratories, and staff working hours. From a practical standpoint, sampling may not be feasible on weekends, and analysis for some samples with

short hold times may not be practical for events sampled late on Fridays if laboratories are closed on weekends.

4) Specify the smallest unit on which decisions or estimates will be made. For BMP projects seeking to compare inflow and outflow water quality, the smallest unit for decision making is often whether or not there is a statistically significant difference between inflow and outflow mean or median concentrations at a given level of significance. For other monitoring efforts, it may be desirable to determine if a reduction in concentrations or mass is greater than or equal to a specific value.

2.4.2 Outputs from Step 4

Step 4 outputs include:

1) Definition of target populations and description of project spatial boundaries. This should include a site map and facility map showing relative physical characteristics related to the tributary watershed and the BMP. Sampling locations should be specified and parameters for each sampling/measurement location identified.

2) Definition of sampling unit. For water quality parameters, a table or list should be assembled with minimum sample volume requirements for laboratory analyses. Sample volume requirements to satisfy QA/QC requirements for replicate analysis, splits, spikes, etc. should also be considered to determine the sample volume required. A “go/no-go” rule should be developed for guidance when storms are partially sampled (i.e., not all stations function as intended, resulting in missed samples).

3) Data collection timeframe, temporal characteristics of monitoring plan and practical constraints. Allowing a reasonable time allotment for preparing a QAPP, a start date for monitoring should be established. The duration of monitoring should be established and constraints on the ability to collect or analyze samples should be identified (i.e., non-daylight hours, weekends, and/or holidays).

4) Scale for decision making or estimation. Based on the questions developed in Step 2, methods for analysis of data should be identified and levels of rejection or acceptance should be established. For statistical testing methods, this will involve specifying the mean hypothesized difference, statistical significance level and potentially other inputs to statistical methods.

2.5 Step 5. Develop the Analytical Approach

This step establishes how data that are collected will be analyzed to provide answers to the questions formulated in Step 2. This step should involve specification of relevant population parameters for analysis and statistical method that will be applied.

2.5.1 Key Activities for Step 5

Key activities for Step 5 include the following:

1) Specify appropriate population parameters for making decisions or estimates. Population parameters for decisions and estimates will be driven by the goals developed in Step 2. For many BMP effectiveness studies, the parameter of interest may be the mean or median EMC, reflecting representative performance of the BMP over the study period. Mean and median population values lend themselves well to statistical hypothesis testing. Other population parameters may be appropriate for other goals. For example, if the purpose of monitoring is for compliance with a discharge permit, the population parameter of interest could be the maximum measured concentration, the maximum EMC, a weekly or monthly average of measured parameters, etc. Exhibit 2-6 presents some of the more common population parameters for water quality monitoring.

2) For decision problems, choose an "action level" and generate "if…then…else" decision rule. Decision problems in BMP effectiveness monitoring usually arise from the need to meet a specified discharge concentration or loading rate or from the need to demonstrate a certain level of effectiveness. Typically "action levels" will arise from some form of regulation or agreement. For an example, an "action level" specified in a watershed management plan for a southeastern drinking water supply watershed is an allowable total phosphorus load of 0.20 lb/ac/year from development that occurs in the watershed. Residential developers in this watershed are required to adopt "conservation design" methods and LID BMPs and to conduct stormwater monitoring to demonstrate compliance with this loading criterion. In this case, an "action level" would be when data indicate that the annual loading of total phosphorus has exceeded 0.20 lb/ac/yr at an outfall from the development. An example of an "if…then…else" statement for this situation could be:

   > *If the annual total phosphorus loading at the outfall from the developed area is greater than 0.20 lb/ac/year, BMP maintenance and potential need for additional BMPs must be evaluated by the developer to reduce total phosphorus, else the total phosphorus loading from the development shall be deemed acceptable.*

3) For estimation problems, specify the estimator and estimation procedures. For assessment of BMP effectiveness, estimators may include population concentration and/or loading characteristics. For example, runoff flow measurements and flow-weighted composite samples could be used to estimate the mean concentration and annual loading of a pollutant in runoff from a specific land use. A mass balance is another example of an estimation problem, where estimators (based on flow and concentration data, typically) are loadings to and from the facility.

Exhibit 2-6. Examples of Population Parameters and Their Applicability to a Decision or Estimation Problem (Adapted from EPA 2006)

Parameter	Definition	Example of Use
Mean (arithmetic or geometric)	Average	Central tendency: Comparison of middle part of population to Action Level. Appropriate for chemicals that could cause cancer after a long-term chronic exposure. Use of the mean and the total amount of media (e.g., mass of soil or water) allows a planning team to estimate the total amount of a contaminant contained in the soil or water body. The arithmetic mean is greatly influenced by extremes in the contaminant distribution. Thus, for skewed distributions with long right tails, the geometric mean may be more relevant than the arithmetic mean. Neither may be useful, however, if a large proportion of values are below the detection limit. Although the arithmetic mean may not represent a good measure of central tendency in a skewed distribution, it remains an important statistic because the large values may represent the bulk of loads. For mass analysis, event mean concentrations (EMCs) from individual events can be flow-weighted based on event runoff volume to develop a flow-weighted average EMC for a series of events over a time period.
Median	Middle observation of distribution; 50th percentile; half of data is above and half is below	Better estimate of central tendency for a population that is highly skewed (nonsymmetrical). Also may be preferred if the population contains many values that are less than the measurement detection limit. The median is not a good choice if more than 50% of the population is less than the detection limit because a true median does not exist in this case (e.g., it would be reported as less than the detection limit). The median is not influenced by the extremes of the contaminant distribution.
Percentile	Specifies percent of sample that is below the given value; e.g., the 80th percentile should be chosen if you are interested in the value that is greater than 80% of the population.	For cases where only a small portion of the population can be allowed to exceed the Action Level. Sometimes selected if the decision rule is being developed for a chemical that can cause acute health effects. Also useful when a large part of the population contains values less than the detection limit. Often requires larger sample sizes than mean or median.

2.5.2 Outputs from Step 5

1) Identification of relevant population parameters.

2) “If…then…else” decision rule based on “action level” for decision problems.

3) Estimation method for estimation problems.

2.6 Step 6. Specify Performance or Acceptance Criteria

Performance and acceptance criteria for BMP effectiveness studies must be specified to define the level of confidence in analysis. Depending on the type of analysis that will be performed (i.e., statistical hypothesis testing, calculations to estimate parameters of a data population).

2.6.1 Key Activities for Step 6

1) Establish criteria for hypothesis testing. Hypothesis testing involves developing a null hypothesis, or baseline condition and an alternate hypothesis, or alternative condition. As an example, the null hypotheses could be that the mean outflow EMC of total suspended solids is not statistically different from the mean inflow EMC. The alternative hypothesis would be that the total suspended solids mean EMCs are different. When testing these hypotheses, data are analyzed at a specified confidence level, and results are interpreted as to whether to accept or reject the null hypothesis. Statistical hypothesis testing has four possible outcomes, summarized in Exhibit 2-7.

Exhibit 2-7. Potential Outcomes of Statistical Hypothesis Testing (EPA 2006)

Decision You Make by Applying the Statistical Hypothesis Test to the Collected Data	**True Condition (Reality)**	
	Baseline Condition is True	Alternative Condition is True
Decide that the Baseline Condition is True	Correct Decision	***Decision Error (False Acceptance)***
Decide that the Alternative Condition is True	***Decision Error (False Rejection)***	Correct Decision

The purpose of this step is to set upper limits on probabilities of false rejection (Type I) and false acceptance (Type II) decision errors. The probability of a false rejection decision is typically referred to as alpha (α) and is the test's level of significance. The probability of a false acceptance decision is called beta (β). The statistical power of a hypothesis represents the probability of a "true rejection" decision and is equal to 1-β. Criteria should be established for both α and β as a part of this step. Figures showing the relationship between α, β, the magnitude of difference being tested and the number of samples are provided in Appendix C and may be useful for determining reasonable α and β values to use given the sampling and budget resources of the monitoring effort. More detailed discussion of hypothesis testing is provided in Chapter 7.

2) Establish criteria for estimates calculated based on data. For parameters that are estimated based on data collected (for example, mean total phosphorus EMCs in BMP inflows and outflows), it is necessary to quantify the uncertainty of the

estimate. The bias and precision associated with data collected directly impact the level of uncertainty in parameter estimates. Bias and precision (collectively known as accuracy) are two principal attributes of data quality in environmental studies. Bias represents systematic error (i.e., persistent distortion that causes constant errors in a particular direction), while precision represents random error (i.e., error among repeated measures of the same property under identical conditions, but not systematically in the same direction or of the same magnitude) (EPA 2006). Useful techniques for quantifying uncertainty include the following:

- Standard Errors: The standard error depends on factors that include the amount of data available, the underlying distribution, and the variability in the data used to calculate the parameter estimate. A standard error can be expressed in either absolute form (i.e., a single number that accompanies the estimate) or relative to the value of the parameter estimate (i.e., a proportion or percentage of the estimate). As an example of a relative standard error uncertainty criterion, a goal could be that the standard error not exceed 30 percent of the mean total phosphorus EMC estimate (EPA 2006).

- Confidence Intervals: A confidence interval is an interval used to estimate a population parameter from sample data. It is generally composed of two parts, an interval calculated from the data and a confidence level associated with the interval. The confidence interval is generally of the form: point estimate ± margin of error. The point estimate is a single value computed from the sample data (for example, a mean pollutant EMC concentration). To account for the possibility of estimation error, the margin of error is included in the confidence interval to provide a range of possible parameter values. The margin of error is what determines the width of the confidence interval (EPA 2006).

  In addition to the confidence interval, there is a confidence level associated with the interval. A confidence level gives the probability that the interval will capture the population parameter in repeated sampling. The level of confidence is expressed in terms of a percentage (e.g., 95 percent confidence). The larger the percentage, the more confidence that the interval contains the true value of the parameter. Consequently, the higher the confidence level, the wider the interval. Thus there is a trade-off between the confidence level and the interval width (EPA 2006).

- Tolerance Intervals: Tolerance intervals are similar to confidence intervals in that they portray uncertainty in a population parameter; however with tolerance intervals the parameter is a specified proportion of the population distribution. Specifically, tolerance intervals estimate the range that should contain a certain percentage of the values in the population. Similar to the concept of confidence level, researchers cannot be 100 percent confident that that interval will contain the specified proportion, only a certain percentage. There are two different inputs associated with the tolerance interval: a degree of confidence and a percent coverage. An example of tolerance interval could

state 95 percent confidence that 90 percent of the population will fall within the range specified by the tolerance interval (EPA 2006).

- Prediction Intervals: While confidence and tolerance intervals estimate present population characteristics, the prediction interval estimates what future values will be, based upon previously collected data. Just as with confidence and tolerance intervals, prediction intervals incorporate the idea of a confidence level when attempting to determine what future values will be. For example, a goal could be to predict that the next set of samples will fall within a determined range with 99 percent confidence. To calculate prediction limits, estimates of the current population mean and standard deviation are needed. It is also necessary to decide how many sampling periods there will be and how many samples will be collected per sampling period. Once these factors are determined, a prediction interval can be calculated for estimating those future observations. Prediction intervals are always larger than confidence intervals (EPA 2006).

3) Identify conditions for site abandonment. In some cases, site conditions may either change over time or be misunderstood to an extent that the monitoring plan will not result in collection of data that are meaningful for the goals and objectives of the monitoring plan. Additionally, in some cases, a BMP being monitored may fail and no longer be viable for monitoring. For example, ultra clean sites may not yield useful information on the ability of the BMP to improve water quality and other sites may have problematic conditions such as backwater, abnormally high loading conditions, etc.

2.6.2 Outputs from Step 6

For studies that will involve hypothesis testing, the following outputs from Step 6 are necessary:

1) Well-formulated baseline condition statement (null hypothesis) and alternative conditions statement (alternative hypothesis).

2) Evaluation of consequences for false rejection and false acceptance decision errors.

3) Specified levels for α and β.

For studies involving estimation of a parameter from a population of data, outputs should include the following:

1) For a criterion using standard error, the desired maximum standard error, either as an absolute value or a percentage of the estimated parameter, should be specified.

2) For statistical intervals, values to specify include the confidence level that specifies the likelihood that the interval contains the true value of the parameter and the desired maximum width of the interval.

2.7 Step 7. Develop the Plan for Obtaining Data

After completing Steps 1 through 6, much of the information needed to develop a thorough and effective plan for collecting data is available. The objectives and goals have been established, types of measurements and estimates have been identified and criteria have been developed for performance or acceptance of results. The next step is to create a QAPP that describes the details of the sampling and analysis techniques that will be used. The sampling design in the QAPP should seek to maximize information that contributes to satisfying project goals within the budget and personnel constraints of the project. This section provides an overview of the elements of a QAPP. Very detailed guidance is available from EPA in the document *Guidance for Quality Assurance Project Plans* (EPA 2002).

2.7.1 Key Activities for Step 7

Key activities for Step 7 include the following:

1) Identify constraints affecting sampling and analysis design.

2) Describe in detail sampling and analyses that will be used.

3) Evaluate efficiency of sampling and analysis design.

The sampling and analysis design should be based on the DQOs developed through Steps 1 through 6 and the specific objectives and intended use of the data (hypothesis testing, estimation). Practical constraints should be considered in developing the QAPP.

2.7.2 Outputs from Step 7

The primary output from Step 7 is the QAPP. Key elements for a QAPP are categorized into these four groups (EPA 2006):

1) Group A – Project Management: These elements address project management, project history and objectives, and roles and responsibilities of the participants. These elements help ensure that project goals are clearly stated, that all participants understand the project goals and approach, and that the planning process is documented.

2) Group B – Data Generation and Acquisition: These elements cover all aspects of the project design and implementation (including the key parameters to be estimated, the number and type of samples expected, and a description of where, when, and how samples will be collected). They ensure that appropriate methods for sampling, analysis, data handling, and QC activities are employed and documented.

3) Group C – Assessment and Oversight: These elements address activities for assessing the effectiveness of project implementation and associated QA/QC requirements; they help to ensure that the QAPP is implemented as prescribed.

4) Group D – Data Validation and Usability: These elements address QA activities that occur after data collection or generation is complete; they help to ensure that data meet the specified criteria.

Exhibit 2-8 summarizes sections that should be included in the QAPP.

Exhibit 2-8. Elements of a QAPP (Source: EPA 2002)

Group A. Project Management	
A1 Title and Approval Sheet A2 Table of Contents A3 Distribution List A4 Project/Task Organization	A5 Problem Definition and Background A6 Project/Task Description A7 Quality Objectives and Criteria A8 Special Training/ Certifications A9 Documentation and Records
Group B. Data Generation and Acquisition	
B1 Sampling Process Design (Experimental Design) B2 Sampling Methods B3 Sample Handling and Custody B4 Analytical Methods B5 Quality Control	B6 Instrument/Equipment Testing, Inspection, and Maintenance B7 Instrument/Equipment Calibration and Frequency B8 Inspection/Acceptance of Supplies and Consumables B9 Non-direct Measurements B10 Data Management
Group C. Assessment and Oversight	
C1 Assessments and Response Actions	C2 Reports to Management
Group D. Data Validation and Usability	
D1 Data Review, Verification, and Validation D2 Verification and Validation Methods	D3 Reconciliation with User Requirements

2.8 Step 8. Assess Reasonableness of Plan and Refine

Once a monitoring plan (QAPP) has been developed, it is time to revisit the objectives and goals developed through the planning process to be sure that the plan will provide the data desired. Refinements to the plan at this step are common, and it is likely that even more refinements and changes to the plans will be required as implementation of the plan begins in the field and the laboratory and unforeseen practical constraints come into play.

2.8.1 Determine Project Costs and Funding Availability

One of the most important "reasonableness assessments" of a monitoring plan is to develop a detailed budget and compare it with available funding. Costs that must be accounted for include the following:

1) Equipment: Capital costs, installation costs, on-going maintenance, replacement parts, repair, etc.

2) Staff time: Field sampling, sample handling and transport, field trips for "false alarms" when samplers malfunction, coordination between team members and laboratory, data entry, data analysis, report writing (progress updates and final), photographic logs, cleaning equipment/bottles, troubleshooting, administrative support, etc.

3) Expenses: Mileage, copies, telephone, field supplies (waders, gloves, hardware, ice), specialized computer programs for data management/analysis, etc.

4) Laboratory/Analytical: Costs for analysis of samples collected for the suite of parameters, including samples needed for QA/QC such as field blanks and replicates. Commercial laboratories generally will provide price quotations if provided with the total number of samples and the desired parameters (be sure to be specific about desired methods and detection limits, which may be need to be lower than for typical laboratory analysis purposes).

5) Peer Review: It is often desirable to have a third party perform a peer review of monitoring data and interpretive reports to increase credibility of the data collected and conclusions.

6) Contingency: The QAPP should provide sufficient detail to develop a detailed cost estimate; however, given the uncertainty inherent in stormwater monitoring, a contingency should also be included in the budget estimate. For example, damage to equipment from animals, lightning and vandalism is not unusual.

2.8.2 Refine Plan Objectives Based on Budget Limitations

If the calculated budget is greater than the amount of funding available, it will either be necessary to narrow the scope of the monitoring project or obtain additional funding.

2.8.2.1 Prioritizing Goals and Data Collection Efforts

The first step to refining a monitoring plan to reduce the required budget is to revisit the overall objective from Step 1 and the goals developed as a part of Step 2. Representative questions to ask as a part of this assessment include:

- Is it possible to achieve the overall objective by addressing a subset of the questions/goals from Step 2?

- Do some of the goals have overlapping data collection requirements, making that data more "valuable" from the standpoint of addressing multiple goals?

- Are there some types of data that would render other types of data less useful if not collected (for example, if flow data are not collected, the utility of concentration data collected is diminished because a mass balance may not be reasonably calculated)?

- Can the list of parameters analyzed be reduced by using "surrogate" parameters as indicators? For example, could monitoring of total suspended solids effectiveness be used to infer, in a general sense, removal of particulate metals for a given pH range?

- Once monitoring has commenced, can the list of parameters analyzed be reduced based on laboratory results? For example, if dissolved cadmium is not detected at the specified detection limit over the first several storm events in inflows or outflows from a constructed wetland, can it be eliminated as a required parameter? Eliminating repeated analysis of samples that result in non-detects can result in significant budget savings during the implementation phase of a monitoring plan.

- Can some parameters be estimated rather than measured? This is a useful question to explore not only for prioritization of data collection efforts but also for filling in data "gaps" that result from missed samples, equipment malfunction, human error and a variety of other causes. (See Section 2.8.2.2 for a discussion of estimation techniques.)

- Is "breadth" or "depth" more important for the data that are being collected? There is an inverse relationship between the number of parameters that can be measured/tested for and the number of measurements for each parameter that can be collected. Often, sample size requirements to meet DQOs will restrict the number of parameters that can be tested under a given budget. If additional parameters are desired, it may be necessary to scale back the number of samples collected and relax acceptance and performance criteria. Alternatively, the list of parameters analyzed may be curtailed to maintain the desired performance and acceptance criteria.

2.8.2.2 Estimation of Parameters for Plan Refinement to Reduce Costs

When monitoring is constrained due to a limited budget or lack of sampling staff, estimates of water quality parameters, flow, and rainfall can be made using various models and assumptions. The use of modeling to estimate these parameters may limit usability of the data, depending on the validity of the assumptions made, the accuracy of the model itself, and accuracy of the information input into the model. Methods to estimate water quality, flow and rainfall parameters follow.

2.8.2.3 Estimates of Water Quality Parameters

Certain water quality parameters can be estimated by monitoring for related parameters that are simpler or less expensive. These related or surrogate parameters are statistically correlated to the more complicated or expensive parameters. Some common surrogate parameters and represented parameters are:

Surrogate Parameter	Parameter Represented by Surrogate
Turbidity	TSS
E. coli	Pathogens

Chemical Oxygen Demand (COD) Biological Oxygen Demand (BOD)

In addition to monitoring for surrogate parameters at each monitoring site, water quality models can be used to estimate constituent concentrations at monitoring sites using available monitoring data, upstream land use, hydrology, geology, and history to calculate a mass balance for each constituent. Water quality models are a tool for simulating the movement of precipitation and pollutants from the ground surface through pipe and channel networks, storage treatment units, and finally to receiving waters. Both single-event and continuous simulation may be performed on catchments having storm sewers and natural drainage for prediction of flows, stages and pollutant concentrations. Each water quality model has its own unique purpose and simulation characteristics. It is advisable to thoroughly review downloading and data input instructions for each model.

The applicability and usefulness of these models is dependent upon a number of assumptions. The degree of accuracy of these assumptions determines the usefulness of the output data. For example, one assumption could be based on certain types of land use contributing certain constituents to the catchment runoff. The constituents associated with each land use have been well studied by many monitoring programs, but are still highly variable, depending on specific activities on each parcel, history of spills, age of infrastructure, climate, and many other factors. Although modeling of water quality parameters is a useful tool to estimate parameter concentrations, model results should not be interpreted as exact data. Confirmation of water quality model results should be done by monitoring a few storms and/or a few sites, then running the model with the observed conditions as input variables and comparing the results.

2.8.2.4 Estimates of Flow

Under certain conditions, flows entering and leaving a BMP can sometimes be modeled if actual monitoring is prohibitive. Most researchers strongly recommend against “taking shortcuts” in terms of flow monitoring; nonetheless, some basic techniques for estimating flow follow. Flow can be estimated at varying levels of detail using approaches ranging from simple spreadsheets to complex hydraulic simulations of extensive urban drainage networks. The simplest approach is to use the volumetric runoff coefficient approach, which is an empirical relationship that provides an estimate of total volume of runoff based on total volume of rainfall according to the following equation:

Volume of Runoff =

Volume of Rainfall x Volumetric Runoff Coefficient - Depression Storage

This method is usually applied to smaller catchments such as parking lots, rather than entire watershed areas. Where monitoring data have been collected for some calibration period such that an accurate estimate of the volumetric runoff coefficient and depression storage for the watershed can be made, this approach coupled with accurate rainfall data may provide one of the least expensive methods for determining total volume of flow from a watershed on a storm-by-storm basis.

For studies where accurate flow data are fundamental to evaluation of performance of the BMP (i.e., many LID studies), flow estimation techniques are typically not an acceptable alternative to meet the study objectives in cases where flow monitoring is feasible.

2.8.2.5 Estimates of Rainfall

If a nearby rainfall gage is not available, rainfall at the monitoring site can sometimes be approximated using available gages that are located as close as possible and at similar elevation. A network of gages in an area can be analyzed to relate latitude, longitude, and elevation to rainfall. The grid of gages can be expanded and extrapolated to an area lacking any gages, provided that enough rainfall gages exist. This approach has significant limitations, particularly at small sites, and should be used only as a last resort

Because costs of rainfall monitoring equipment are often much less than other types of equipment and because rainfall patterns often vary significantly, it is highly recommended that rainfall monitoring be conducted instead of estimated. This is particularly true for LID-related studies were surface runoff volume reduction is a key measure of performance. Because of localized variability in storm conditions, sites with small tributary areas also generally require a rain gage at the site.

2.9 Conclusion

A systematically and carefully planned monitoring plan is essential to a cost-effective monitoring program that results in appropriate data needed to meet study objectives and goals. BMP monitoring is expensive and care should be taken to ensure that the study design will enable the researcher to draw statistically significant conclusions or meet other objectives such as permit requirements.

2.10 References

Note: This reference list include both publications cited in this chapter and additional resources that may be useful in developing a monitoring plan.

Burton, A. and R. Pitt. 2001. *Stormwater Effects Handbook: A Toolbox for Watershed Managers, Scientists, and Engineers*. Lewis Publishers. http://www.epa.gov/ednnrmrl/publications/books/handbook/index.htm

Caltrans. 2009. Water Quality Guidelines & Protocols BMP Pilot Study Guidance Manual CTSW-RT-06-171.02.1

Chow, V.T. 1954. The lognormal distribution and its engineering applications. *Proc., American Society of Civil Engineers*. 110, 607-628.

Cochran, W.G. 1963. *Sampling Techniques*. Second Edition. John Wiley and Sons, Inc. New York, New York.

Driscoll, E., G. Palhegyi, E. Strecker, and P. Shelley. 1989. *Analysis of Storm Event Characteristics for Selected Rainfall Gauges Throughout the United States*. Draft Rep. Prepared for U.S. Environmental Protection Agency, Woodward-Clyde Consultants, Oakland, CA.

Driscoll, E.D., P.E. Shelly, and E.W. Strecker. 1990. *Pollutant Loadings and Impacts from Stormwater Runoff, Volume III: Analytical Investigation and Research Report*. Federal Highway Administration Final Report FHWA-RD-88-008, 160 p.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. John Wiley and Sons, Inc. New York, New York.

Granato, G. 2009. Personal Communication with Gregory Granato, U.S. Geological Survey.

Harremoes, P. 1988. Stochastic models for estimations of extreme pollution from urban runoff. *Water Resource Bulletin*, 22, 1017-1026.

Hunt, W. 2003. Pollutant Removal Evaluation and Hydraulic Characterization for Bioretention Stormwater Treatment Devices. A Thesis in Agricultural and Biological Engineering. Pennsylvania State University. August.

Jones, J., J. Clary, E. Strecker, M. Quigley. 2008. "15 Reasons You Should Think Twice Before Using Percent Removal to Assess BMP Performance." *Stormwater*. January/February. Vol. 9, No. 1.

Loftis, J.C., L.H. MacDonald, S. Street, H.K. Iyer, and K. Bunte. 2001. *Detecting Cumulative Watershed Effects: The Power of Pairing*. In Review *Journal of Hydrology*.

Pitt, R. 1979. *Demonstration of Nonpoint Pollution Abatement Through Improved Street Cleaning Practices*, EPA-600/2-79-161, U.S. Environmental Protection Agency, Cincinnati, Ohio.

Pitt, R. 2009. Effort for Paired Samples. Unpublished manuscript. March 6.

Pitt, R. and K. Parmer. 1995. *Quality Assurance Project Plan (QAPP) for EPA Sponsored Study on Control of Stormwater Toxicants*. Department of Civil and Environmental Engineering, University of Alabama at Birmingham.

Strecker, E.W. 1994. Constituents and Methods for Assessing BMPs. In *Proceedings of the Engineering Foundation Conference on Storm Water Monitoring*. August 7-12, Crested Butte, CO.

Strecker, E.W., J.M. Kersnar, E.D. Driscoll, and R.R. Horner. 1992. The use of wetlands for controlling storm water pollution. The Terrene Institute, Washington, D.C.

Urbonas, B., and Stahre, P., 1993. *Stormwater: Best management practices and detention for water quality, drainage, and CSO management*, Prentice Hall, Englewood Cliffs, NJ.

Urbonas, B.R. 1994. Parameters to Report with BMP Monitoring Data. In *Proceedings of the Engineering Foundation Conference on Storm Water Monitoring*. August 7-12, Crested Butte, CO.

U.S. Environmental Protection Agency (EPA). 1983. *Results of the Nationwide Urban Runoff Program, volume I – Final Report.* U.S. Environmental Protection Agency, Water Planning Division, Washington D.C.

U.S. Environmental Protection Agency. 2002. *Guidance for Quality Assurance Project Plans*. EPA QA/G-5, USEPA, Office of Environmental Information, Washington, D.C. http://www.epa.gov/quality/qs-docs/g5-final.pdf

U.S. Environmental Protection Agency. 2006. *Guidance on Systematic Planning Using the Data Quality Objectives Process*. EPA QA/G-4, USEPA, Office of Environmental Information, Washington, D.C. http://www.epa.gov/quality/qs-docs/g4-final.pdf

Van Buren M.A., W.E. Watt, and J. Marsalek. 1997. Applications of the Log-normal and Normal Distributions to Stormwater Quality Parameters. *Wat. Res*., Vol. 31, No 1, pp. 95-104.

Watt, W.E., K.W. Lathem, C.R. Neill, T.L. Richards, and J. Rouselle, (Eds). 1989. *Hydrology of Floods in Canada: A Guide to Planning and Design*. National Research Council Canada, Ottawa, Ontario.