

SSFL CDO Expert Panel¹

Robert Gearheart, Ph.D., P.E.
Richard Horner, Ph.D., P.E.
Jonathan Jones, P.E., D.WRE
Michael Josselyn, Ph.D.
Robert Pitt, Ph.D., P.E., BCEE, D.WRE
Michael K. Stenstrom, Ph.D., P.E., BCEE

Sample Collection Methods for Runoff Characterization at Santa Susana Field Laboratory

Executive Summary

This white paper reviews methods for sampling stormwater discharges at the Santa Susana Field Laboratory (SSFL). Based on the scientific and engineering literature, flow-weighted composite sampling is identified as the most suitable method for monitoring compliance with the SSFL permit limits. This paper was prepared at the request of the Los Angeles Regional Water Quality Control Board (Regional Board).

The SSFL has a National Pollutant Discharge Elimination System permit for stormwater discharges (NPDES permit number R4-2007-0055) that is administered by the Regional Board. The current permit requires collection of grab samples for runoff characterization to assess permit compliance. The permit states that “sampling shall be during the first hour of discharge or at the first safe opportunity” (State of California Los Angeles Regional Water Quality Control Board 2007 d, p T-7, footnote 1, 2007).

The underlying basis for the requirement of sampling during the first hour of discharge is a general assumption by the Regional Board’s stormwater regulatory program that the “first flush” concept is applicable for all sites. The first flush refers to the beginning of a runoff event when concentrations of contaminants are theoretically highest, such as when the initial runoff from a rainstorm flushes a street or parking lot.

Since the 1990s, multiple researchers have found that a first flush is not always observed in all watersheds nor is it observed for all pollutants. The first flush phenomenon is most pronounced in watersheds that are smaller or that have a large fraction of impervious area. In larger watersheds or in those with less impervious area (i.e., with a greater proportion of open space), the first flush effect is less likely to be exhibited.

Many of the SSFL watersheds are large and primarily open space (over 90% of the SSFL is undeveloped area), and the ongoing removal of pavement and demolition of buildings at the site is

¹ The Expert Panel members are acting as private consultants in order to assist the Regional Board and The Boeing Company develop and implement methods to meet the requirements of Cease and Desist Order R4-2007-0056, dated November 1, 2007. Their opinions and directives are not the opinions and directives of their respective employers.

causing a reduction in the impervious area over time. In addition, the implementation of Engineered Natural Treatment Systems (ENTS) throughout watersheds 008 and 009 will further dampen peak flow rates being discharged from the site. These combined factors suggest that the first-flush pattern of pollutant discharge is not an appropriate assumption with respect to the SSFL; hence, a sampling methodology based on grab samples collected during the first hour of stormwater discharge is likely to not capture the maximum concentration of a contaminant or accurately represent the pollutant load over the duration of the event.

Within the realm of what is physically practical, grab sampling is a poor method for characterization of the event mean concentration (EMC) and mass loading of an event. The costs of processing and analyzing the complete analytical suite for the necessarily large number of grab samples per location per event make such an approach uneconomical for the information gained. The best method for characterization of the “true” EMC is flow-weighted composite sampling using automatic sampling equipment interfaced with flow monitoring equipment to collect many subsamples over the course of the hydrograph. Accuracy of the estimate of the “true” EMC increases as more subsamples are collected.

A wide body of scientific and engineering literature indicates that flow-based composite sampling is the best method to represent concentrations of contaminants over the duration of a stormwater discharge event. A composite sampling methodology reduces the risk of missing either the low or the high concentrations of a specific constituent in the runoff. The SSFL NPDES permit specifies compliance with the permit limits based on the daily maximum concentration, not the instantaneous maximum, of each constituent. Consequently, the sampling methodology used should provide the best known approach for capturing a representative sample of the entire runoff event. A grab sample will not, in all probability, provide a representative sample of the runoff from a storm event. However, due to holding time and sampling equipment limitations, some parameters such as oil/grease and VOCs will still require manual grab sampling, which can still be conducted at the start of the event or when the site is accessible for events that begin during non-daylight hours.

Collection of a series of individual, discrete samples over the course of a storm event with automated sampling equipment has been suggested by some stakeholders as an alternative to collecting a single composite sample composed of multiple samples collected during runoff for a storm. However, there are practical constraints that make this infeasible except on a very limited basis and for an extremely abbreviated list of parameters. Analysis of discrete samples over a hydrograph for the suite of regulated parameters at all the SSFL NPDES outfalls, versus composite samples, would require collection and handling of large sample volumes of runoff—a dedicated cooler truck would be needed to transport samples to the laboratory for analysis.

There are also physical limitations of automated sampling equipment. A typical automated sampling unit is capable of collecting up to 24 discrete 575 mL samples (if glass bottles are required, the container volumes are only 300 mL) (American Sigma 2008). For laboratory analysis of metals alone, Standard Methods for the Examination of Water and Wastewater (American Public Health Association 1995) specifies a minimum sample volume of 500 mL, meaning there would be an inadequate sample volume remaining for analysis of other parameters, such as dioxin, for multiple discrete samples over the course of a runoff event. While this type of discrete analysis has been conducted in a research setting for specific analytes, including many of the studies cited in this paper that advocate flow-weighted composite sampling, in the authors’ experience, this level of analysis is unheard of in the context of permit compliance.

Based on the sampling objective for the SSFL, combined with the large body of evidence in the scientific and engineering literature and the professional judgment of the SSFL Expert Stormwater Panel, the Panel recommends that a flow-weighted composite sampling approach be used to best characterize the “true” EMC for each constituent to assess compliance with the NPDES stormwater permit.

Introduction

This white paper provides a review of the engineering and scientific literature related to stormwater runoff sampling methods for discharge compliance monitoring at the Santa Susana Field Laboratory (SSFL) in Southern California. It was prepared based upon the request of the Los Angeles Regional Board Water Quality Control Board (Regional Board) at the April 3 public hearing after Dr. Michael Stenstrom of UCLA provided an update on the Expert Panel’s SSFL efforts.

Specific topics addressed in this white paper include:

- Representativeness of grab sampling,
- Sampling to characterize the “first flush” and
- Relative strengths and weaknesses of grab and composite sampling techniques.

This white paper includes a review of many documents that define the state of the practice for stormwater sampling techniques, which overwhelmingly favor flow-weighted composite sampling over grab sampling for characterization of the storm Event Mean Concentration (EMC). The EMC is most meaningful for accurately determining pollutant loads from a site, and is most representative of average pollutant concentrations over an entire runoff event.

The SSFL site has a National Pollutant Discharge Elimination System (NPDES permit number R4-2007-0055) permit for stormwater discharges that is administered by the Los Angeles Regional Water Quality Control Board (LARWQCB). As it currently stands, this permit requires collection of grab samples for runoff characterization to assess permit compliance. Currently, the permit states that “sampling shall be during the first hour of discharge or at the first safe opportunity” (State of California Los Angeles Regional Water Quality Control Board 2007 d, p T-7, footnote 1, 2007). Given the fact that many of the storms at the SSFL site follow typical southern California weather patterns, occurring in the winter months and beginning in non-daylight hours, from practical and safety standpoints, this permit requirement for sample collection is satisfied by collecting manual grab samples from outfalls the mornings following the start of runoff events when daylight allows safe sample collection, typically hours after the start of rainfall and runoff. For shorter duration storm events, this can result in sampling of the tail end of the hydrograph rather than a “first flush.” Even for longer, multi-day events, this practical constraint results in the inability to collect a sample during peak flow (which is generally sharp and brief, therefore catching this period with a manual grab sample is unlikely) and/or the “first flush” of pollutants (i.e., beginning of runoff when concentrations are theoretically highest) that is hypothesized to occur in the initial runoff leaving the site.

Most storms in southern California begin in the early evening after dark or during very early morning hours before sunrise during the winter when daytime temperatures drop. With limited

access and unsafe streamflow conditions at most of the outfall monitoring locations, manual sampling can't safely occur until daylight. Appendix A shows that about 2/3 of the rainfall in Southern California is expected to initiate during dark periods of the day. Limiting samples to daylight periods would only allow a very few events to be sampled during their first hour of runoff each year. Therefore, for most events, samples are collected during the tail end of short overnight storms or otherwise are just unlikely to be timed with either the start of the storm or the peak discharge period. This could of course be replaced with an automatic discrete sampler programmed to occur at the start of runoff, but then there's the issue of whether the "first flush" concept, as discussed in Appendix B, would be expected for a wide range of parameters from very pervious watersheds.

The Regional Board's underlying basis for this sample timing requirement is because of a general assumption inherent to their stormwater regulatory program that the "first flush" concept holds for all sites. As shown in Appendix B, this is not a universally applicable concept and may not be valid for this site. This is especially true for large sites that have little development such as at several of the SSFL outfalls. With the diffused Engineered Natural Treatment Systems (ENTS) and large detention facilities throughout watersheds 008 and 009, the hydrographs will be further dampened and delayed, while the ENTS treatment will dramatically reduce the peak concentrations that do occur. Therefore, the blended outfall discharges will be even more consistent and less likely to have peak contaminant concentrations near the beginning of the runoff events.

Stormwater discharges from the site are subject to numeric limits for a large list of parameters at more than a dozen outfalls (State of California LAWQCB 2007 a, b and c). The Expert Panel is focusing primarily on two outfalls (Outfalls 008 and 009) for the design of ENTS; however, the Panel believes that sampling recommendations would be relevant for consideration at all of the outfalls on the site with numeric limits for stormwater runoff quality.

The site is undergoing clean up and restoration to a more natural, open space condition, which is currently scheduled for completion in 2017. The proposed plan for treating stormwater runoff, until the scheduled site cleanup is completed in 2017 for the Outfall 008 and 009 watersheds, relies on ENTS including bioswales and bioretention practices. These systems are designed to fully treat at least the 1-year event (some ENTS will fully treat up to between the 2- and 5-year events) and provide significant partial treatment for larger events (i.e. full treatment of the design volume, plus partial treatment for additional volume). The general basis for the ENTS-based approach to water quality management at these watersheds is the Cease and Desist Order as well as direction from the Regional Board (State of California Los Angeles Regional Water Quality Control Board 2007a, Finding 43).

The issue of grab versus composite sampling has previously been investigated at SSFL in a study conducted in 2006. Boeing reported results comparing grab and composite water quality samples collected below Outfall 11 during 2006 in the report titled *Results of the Evaluation of the Grab Versus Flow Weighted Samples and Evaluation of Filtered Versus Unfiltered Samples for Radionuclides* (The Boeing Company 2006). The study that generated this data was collected based on a request from the Regional Water Quality Control Board pursuant to Section 13267 of the California Water Code. While this study collected valuable data that serve as a starting point for quantifying some of the issues identified in this white paper, the limited number of storms evaluated, coupled with a large number of sample results below the method detection limits, limited

the conclusions that could be drawn from the data. The data collected showed considerable variability and did not exhibit a clear trend in comparison of grab and composite sample results. In some instances the composite sample result exceeded the grab sample result and visa versa. One of the purposes of this white paper is to further the understanding of the grab versus composite sampling topic as it relates to the SSFL site based on broader data collection efforts presented in the literature.

Based on a review of the scientific and engineering literature, there are many findings related to sampling methods that are relevant for determining the appropriate types of stormwater samples to collect to characterize runoff conditions. These sample collection methods include: manual grab sampling, manual or automated collection of multiple discrete samples during an event, time or flow-weighted composite sampling using automated samplers, or combinations. Findings and supporting literature are summarized in the following sections. The general consensus from this literature is that flow-weighted composite sampling using automated sampling equipment is the “gold standard” for determining representative pollutant concentrations and loads and for assessing performance of water quality facilities such as the ENTS.

Figure 1 provides a visual comparison of manual grab sampling and automated composite sampling “windows” that would be applicable for SSFL. The data used to illustrate the hydrograph and pollutograph were taken from the Southern California Coastal Water Research Project (SCCWRP) report *Assessment of Water Quality Concentrations and Loads from Natural Landscapes* (Stein and Yoon 2007). The “window” for manual grab sampling is assumed to be up to 15 hours. For storms with runoff beginning in daylight hours, a manual grab sample would be collected within the first hour of discharge. For storms beginning in non-daylight hours there could be up to 15 hours before a sample is collected (event starting after 5 p.m. would be sampled the next day at 8 a.m.). For composite sample collection using automated equipment, however, aliquots could be collected over the duration of the hydrograph, resulting in a much more representative sample. Another significant point illustrated by Figure 1 is a delayed peak concentration (outside of the potential 15-hour grab sampling window).

Manual Grab Samples for Characterization of Event Mean Concentration

The use of grab samples to characterize runoff concentrations has been a topic of considerable research in southern California as well as the rest of the United States. Research efforts have typically sought to compare EMC results for parameters determined from a more limited series of grab samples versus the “true” EMC, approximated by very frequent sampling of small volumes of runoff over the duration of the hydrograph. Stenstrom and Kayhanian (2005) used a stochastic regression model to evaluate the accuracy of various grab sampling strategies to characterize the “true” EMC. They found that with 10 grab samples collected over the hydrograph, estimates of the “true” EMC were poor, with median errors of 40% for randomly timed grabs and 23% for samples collected at equal flow volumes. Substantial errors are still present even for 20 samples, with median errors of 30% and 16% for time- and flow-proportioned grab samples, respectively. When 100 or more individual samples are collected and combined to create a composite sample (easily accomplished with many automatic samplers but impractical for manual grab sampling), median errors for both time- and flow-proportioned samples can theoretically be less than 10%, although sampling errors and biases generally would not allow such low median errors in a practical sense. Grab sampling, within the realm of what is physically practical, is a poor method for

characterization of the EMC and mass loading of an event (BMP Database Project Team 2002, FHWA 2001, Maestre et al. 2004, Stenstrom and Kayhanian 2005).

Grab Samples for Characterization of First Flush or Maximum Event Concentration

While it is clear in the literature that grab samples are poor predictors of EMCs, individual grab samples have often been used in an attempt to characterize the “first flush” or to try to capture a maximum concentration in runoff for an event. The concept of a “first flush” dates to the early 1970’s when runoff sampling methods required collection and analysis of multiple discrete samples over a storm hydrograph. Data from stormwater monitoring in urban areas demonstrated that pollutant concentrations for some pollutants (those associated with particulates) tended to be higher at the beginning of runoff (Scheuler and Holland 2000). More recent research has refined the understanding of the first flush in terms of mass, concentration, seasonal versus intra-event and other factors (Stenstrom and Kayhanian 2005). Multiple researchers since the 1990s have found that a first flush is not always observed in all watersheds nor is it observed for all pollutants. Studies and articles including Chang et al. (1990), Brown et al. (1995), Hager (2001), Maestre et al. (2004), Stenstrom and Kayhanian (2005) and New South Wales Australia (2008) have identified factors affecting the first flush including the following:

- The first flush phenomenon is most pronounced in watersheds that have high levels of imperviousness. As one example, Chang et al. (1990) found that the first flush effect was weak for sites with imperviousness ranging from 5 to 30% in a study of seven urban sites with over 160 storm events. Similarly, analysis of data from the National Stormwater Quality Database (NSDQ) (Pitt et al. 2004) that examined fourteen parameters, including total suspended solids, oxygen demands, total dissolved solids, nutrients and metals for an open space land use, did not reveal a first flush for any of these parameters (Maestre et al. 2004)
- Larger watersheds (typically > 400 acres) are less likely than small watersheds to exhibit a first flush, although the travel time generally is a better predictor of the first flush effect than size since this effect is generally a result of time lags of smaller first flushes from individual smaller portions of the watershed (Stenstrom and Kayhanian 2005).
- Some pollutants are more likely to exhibit a first flush than others. Dissolved pollutants and bacteria are generally less likely to exhibit a first flush (Scheuler and Holland 2000, Hager 2001, and Maestre et al. 2004). Han et al. (2006) found that even for highly impervious areas, runoff exhibited a weak first flush for ionic pollutants such as nitrate and nitrite. In some cases, dissolved constituents such as dissolved copper have been observed to sometimes increase during the course of the storm (City of Portland 1996). In addition, Herricks found that the end of storms is often where aquatic toxicity is highest (Herricks and et al. 1997). This could be due to a combination of some constituents being more concentrated at the end of storm event as well as being more bioavailable. For example, total hardness, which affects bioavailability (as recognized in EPA’s Aquatic Life Criteria) is often lower during the end of storms and therefore EPA’s calculated criteria levels would also be lower during the end of the storm.

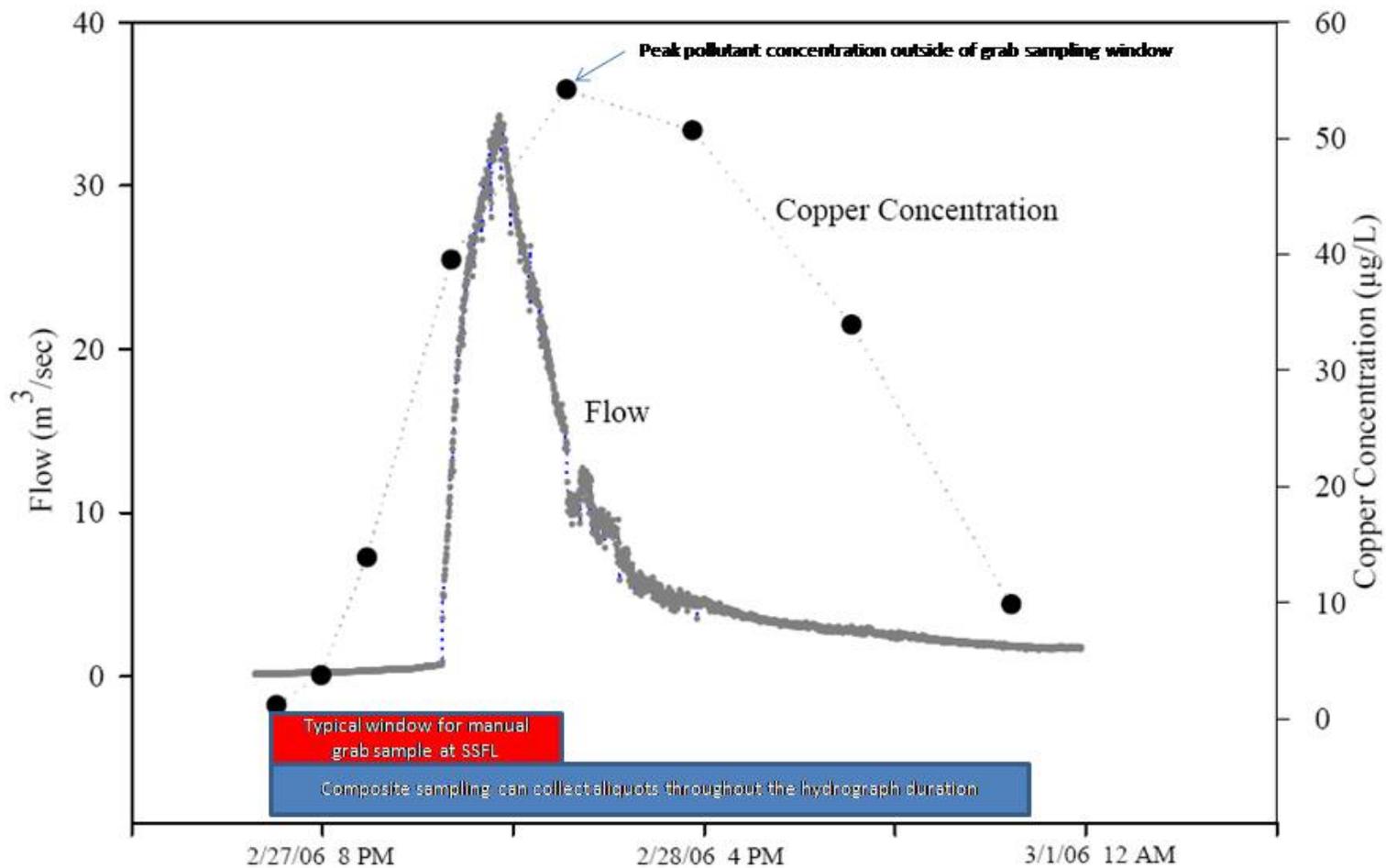


Figure 1. Hydrograph and Pollutograph with Sampling Windows for Manual Grab and Composite Sampling (Piru Creek Data, Southern California February 27 – March 1, 2006 from SCCWRP [Stein and Yoon 2007])

- Transport of available pollutants is another factor influencing the first flush (New South Wales Australia 2008). If pollutants are in locations far removed from the discharge point or if features within the watershed, such as treatment mechanisms, delay or reduce the delivery of pollutants to the outfall, a first flush may not be observed and maximum concentrations may occur later in the event.

Many of the SSFL watersheds are large (especially 001, 002, 008, 009, 011, and 018) and primarily comprised of open space (over 90% of the SSFL is undeveloped area). Furthermore, the continuing demolition of pavement and buildings will result in reduced imperviousness throughout the site over time, and the implementation of ENTS throughout watersheds 008 and 009 will result in dampened/delayed peak flows and well-mixed discharges. These combined factors strongly indicate that the first flush pollutant discharge pattern is a poor assumption for stormwater monitoring at the site, and that flow-based composite sampling will best represent concentrations over the course of the stormwater discharge events, while reducing the risk for missing both low and high concentrations as grab sampling likely would.

It is actually more likely that, after the ENTS are constructed, there would not be an observable consistent first flush in both watersheds 008 and 009. In fact, a sample collected at the beginning of an event might often be “cleaner” than those collected later in the event, particularly for larger events. Further discussion of these issues are presented in Appendix B.

Collection of a series of discrete samples has been suggested by some stakeholders; however, there are practical constraints that make this infeasible except on a very limited basis and for an extremely abbreviated list of parameters. Analysis of discrete samples over a hydrograph for the suite of regulated parameters at SSFL would require collection and handling of large volumes of runoff—a dedicated cooler truck would be needed to transport samples of sufficient volumes to the laboratory for analysis. There are physical limitations of automated sampling equipment as well. A typical automated sampling set up is capable of collecting up to 24, discrete 575 mL samples (if glass bottles are required the container volumes are only 300 mL) (American Sigma 2008). Standard Methods for the Examination of Water and Wastewater (American Public Health Association 1995) specifies a minimum sample volume of 500 mL for metals alone. This type of discrete analysis has been conducted in a research setting, including many of the studies cited in this paper that advocate for flow-weighted composite sampling; however this level of analysis is unheard of in the context of permit compliance.

Standard of Practice – Flow-Weighted Composite Sampling

Flow-weighted composite sampling to characterize runoff water quality is recommended or required by many of the leading agencies across the country. The following are examples of guidance manuals that recommend or require this type of sampling:

- National Research Council (NRC), Committee on Reducing Stormwater Discharge Contributions to Water Pollution, *Urban Stormwater Management in the United States* (NRC 2008). This recently published report states a strong preference for flow weighted composite sampling as opposed to grab sampling:

Continuous, flow-weighted sampling methods should replace the traditional collection of stormwater data using grab samples. Data obtained from too few grab samples are highly variable, particularly for industrial monitoring programs, and subject to greater uncertainty because of experimenter error and poor data-collection practices. In order to use stormwater data for decision making in a scientifically defensible fashion, grab sampling should be abandoned as a credible stormwater sampling approach for virtually all applications. It should be replaced by more accurate and frequent continuous sampling methods that are flow weighted. Flow-weighted composite monitoring should continue for the duration of the rain event. Emerging sensor systems that provide high temporal resolution and real-time estimates for specific pollutants should be further investigated, with the aim of providing lower costs and more extensive monitoring systems to sample both streamflow and constituent loads.

- Southern California Stormwater Monitoring Coalition (SMC), *Model Monitoring Program for Municipal Separate Storm Sewer Systems in Southern California* (SMC 2004). This document states that “flow compositing was the most efficient sampling approach to achieve a given degree of accuracy and precision.” This document was developed by the SMC Model Monitoring Technical Committee, which comprised Regional Water Quality Control Boards (Los Angeles, Santa Ana and San Diego), Municipal Permittees (Counties of Ventura, Los Angeles, San Bernardino, Riverside, Orange and San Diego), Heal the Bay and Southern California Coastal Water Research Project (SCCWRP).
- Washington State Department of Ecology, *How to Do Stormwater Sampling: A Guide for Industrial Facilities*. This document acknowledges that grab samples are collected in many cases, typically due to the cheaper cost, but recognizes the benefits of automated flow-weighted composite samples (Washington State Department of Ecology 2002).
- Federal Highway Administration, *Guidance Manual for Monitoring Highway Runoff (FHWA 2001)*. This document compares the advantages and disadvantages of grab and flow-weighted composite sampling techniques and notes that “a small number of samples are not likely to provide a reliable indication of stormwater quality at a given site.”
- International Stormwater Best Management Practices Database (BMP Database), *Urban Stormwater BMP Performance Monitoring, A Guidance Manual for Meeting the National Stormwater BMP Database Requirements*. The manual was prepared by the BMP Database Project Team in cooperation with the United States Environmental Protection Agency (USEPA) and the American Society of Civil Engineers (ASCE). (BMP Database Project Team 2002). It recommends that flow-weighted composited samples be collected to assess BMP performance and the BMP Database protocols require flow-weighted composite sample collection for those parameters that are appropriate.
- Technology Acceptance Reciprocity Partnership (TARP), *Protocol for Stormwater Best Management Practice Demonstrations*. A composite sampling protocol has been endorsed

by California, Maryland, Massachusetts, New Jersey, Pennsylvania, and Virginia (TARP 2001).

- USEPA, *BMP Performance Webcast*. The webcast discusses the limitations of grab sampling and advocates for flow-weighted composite sampling for characterization of stormwater runoff water quality (USEPA 2008).

Comparative Evaluation of Sampling Methods for SSFL

Although the discussion above focuses on the relative advantages and disadvantages of composite versus grab sampling, capabilities of automatic sampling equipment permit hybrid options where both a first flush grab sample and a flow-weighted composite sample can be collected, as further described in Appendix C. Samplers can be programmed, with various bottle configurations, to collect an initial discrete sample and then to collect subsamples at equal flow increments to create a composite sample (Teledyne ISCO 2004). This leads to a number of potential sampling methods that could potentially be used at SSFL, including the following:

- Option 1—Manual collection of first flush grab sample.
- Option 2—Manual collection of multiple grab samples over course of event.
- Option 3—Automatic sample collection of discrete samples (typically up to 24 discrete samples of reduced volume can be collected at each station, with each automated sampler).
- Option 4—Automatic sample collection of flow-weighted composite sample.
- Option 5—Automatic sample collection of first flush discrete sample and flow-weighted composite sample.

Advantages and disadvantages of each of these options are summarized in Table 1.

Table 1. Comparison of Potential Sampling Options for SSFL (Based on FHWA 2001, Maestre et al. 2004, Teledyne ISCO 2004 with additions)

Option 1. Manual Single First Flush Grab Sample	
Advantages	Disadvantages
<ul style="list-style-type: none"> • Is suitable for any parameter. • Not necessary to sample entire duration of storm. • Easy to move from one location to another due to minimal equipment. • Low equipment and analytical cost. 	<ul style="list-style-type: none"> • Peak pollutant concentrations likely do not coincide with a first flush and may occur at different times in an event for different parameters. • In some watersheds, peak pollutant concentrations may occur during peak flows, which could occur at any time during the storm. No guarantee that grab sample is “worst case.” • Depending on the distance from the pollutant source to the monitoring point, the pollutant may not arrive during the first flush. • Not suitable for calculation of mass loading. • First flush sampling of storms that begin at night, on weekends, holidays, etc. may not be possible. • Safety concerns for sampling staff working outside under storm conditions. • Storm patterns at SSFL are that storms usually start at night and due to health and safety concerns, samples are typically not collected until the next morning (see Appendix A). • Sampling personnel usually do not arrive at the sampling location at the beginning of the event due to delayed and inaccurate indications of rainfall at the site, and the time needed to drive to the sampling station.

Option 2. Manual Grab Samples Over Course of Event	
Advantages	Disadvantages
<ul style="list-style-type: none"> • Less likely to omit a source due to runoff quality variations during the storm. • Is suitable for any parameter. • Potential to manually create flow-weighted composite sample from individual grab samples if flow measurements are available. Could analyze both discrete and composite samples. • Easy to move from one location to another due to minimal equipment. 	<ul style="list-style-type: none"> • Labor intensive. • Practical limitations on number of samples that can be collected to create a composite may lead to significant median errors relative to the “true” EMC. Very frequent sampling over hydrograph at multiple locations is possible but is very costly in terms of staff time and may involve complicated logistics, increased possibility of sampling “errors.” • Moderate to high cost depending on number of grab samples collected and analyzed by discrete analysis. If composite sample is created and analyzed as alternative to discrete samples, analytical costs may be lower. High labor costs. • Access to site over duration of storm event may not be feasible due to safety and other considerations (night time storms, weekends, holidays, etc.) • Safety concerns for sampling staff working outside under storm conditions, particularly during the night. • Delays in reaching site often result in missing first part of runoff period that may be important, especially if the rising hydrograph time period is short.

Option 3. Automatic Sample Collection of Discrete Samples	
Advantages	Disadvantages
<ul style="list-style-type: none"> • Less likely to omit a source due to runoff quality variations during the storm. • May be suitable for more parameters than flow weighted composite sampling techniques since discrete analysis is possible. • Potential to manually create flow-weighted composite sample from discrete samples, assuming automatic sampling equipment capable of interfacing with flow monitors. Could analyze both discrete and composite samples. • Automatic sampling equipment can operate during periods when staff access to site is not possible. • Less labor intensive than manual collection of discrete samples over the hydrograph. 	<ul style="list-style-type: none"> • Labor intensive to manually create composite sample from discrete samples. Manual compositing introduces potential errors. • Limitations on number of discrete samples that can be collected to create a composite likely to lead to significant median errors relative to the “true” EMC (unless more than one automated sampler is deployed at a site) • Moderate to very high cost depending on number of grab samples collected and analyzed individually. • Potential for automatic sampling equipment malfunction. • Moderate cost for equipment set up and maintenance. • Cannot collect true duplicate samples, unless multiple samplers deployed. • May not work for parameters with short holding times.

Option 4. Automatic Sample Collection of Flow-Weighted Composite Samples	
Advantages	Disadvantages
<ul style="list-style-type: none"> • Less likely to omit a source due to runoff quality variations during the storm. • Results provide a better indication of the relative importance of a source than a single grab “snapshot.” • Sampler can be programmed to collect many sub-samples to create composite that is representative of “true” EMC with a low median error. • Can provide EMC that enables calculation of mass loading (especially relevant for Total Maximum Daily Load [TMDL] based on mass). • Equipment can be triggered and collect samples at any time of the day or night. • Reduced health and safety risks for sampling crew. 	<ul style="list-style-type: none"> • Moderate cost for equipment, set up, and maintenance. • Equipment errors and malfunctions can occur. • Cannot collect true duplicate samples. • May not work for short holding times or for parameters that cannot be composited in a representative fashion. • The sample is an average during the event and therefore the maximum concentration is not identified

Option 5. Automatic Sample Collection of First Flush Discrete Sample and Flow Weighted Composite Samples	
Advantages	Disadvantages
<ul style="list-style-type: none"> • Combines advantages of Option 1 and Option 4 	<ul style="list-style-type: none"> • Peak pollutant concentrations do not always coincide with a first flush and may occur at different times in an event for different parameters. • In some basins for some pollutants peak concentrations may occur during peak flows which could occur at any time during the storm. No guarantee that initial discrete sample is “worst case.” • Depending on the distance from the pollutant source to the monitoring point, the pollutant may not arrive during the first flush. • See disadvantages for 1 and 4. Double the analytical cost of Option 4.

Conclusions

Given the above considerations from the scientific and engineering literature, the SSFL Expert Stormwater Panel draws several conclusions:

1. Collection of a single first flush grab sample would not likely be sufficient to characterize the maximum event concentration for runoff events for the suite of regulated parameters at the SSFL site. It is questionable for many parameters whether or not an event first flush effect would occur, particularly after the ENTS are installed. One first flush sample would also not represent the complete storm conditions that are needed to meet the maximum daily concentration limit as specified by the permit limits (see Appendix B).
2. The collection of a large number of manual grab samples or automatically-collected discrete samples is needed to adequately characterize the “true” EMC for a runoff event. The Panel

believes that the problems associated with handling such a large number of samples in the short time necessary before analyses would result in unacceptable processing errors and times delays. It would also be impossible to collect a sufficient amount of sample for many discrete intervals for the complete analytical list using any available sampling equipment. The costs of processing and analyzing the complete analytical suite for the necessarily large number of samples per location per event would be uneconomical for the information gained and may establish an unreasonable precedent.

3. As noted in Appendix B, daily maximum concentrations are specified in the permit, not maximum instantaneous concentrations. These daily maximum values are best determined based on the event-mean-concentration (EMC) of the event. The best method for characterization of the “true” EMC is flow-weighted composite sampling using automatic sampling equipment interfaced with flow monitoring equipment to collect many subsamples over the course of the hydrograph. Accuracy of the estimate of the “true” EMC increases as more subsamples are collected.
4. It may be feasible given the capabilities of automatic sampling equipment to collect both a first flush discrete sample and a flow-weighted composite sample (Option 5), although it still would be unlikely that the first flush discrete sample would be representative of the maximum event concentration for all regulated parameters.

The SSFL Expert Stormwater Panel recommends flow-weighted composite sampling for characterization of the “true” EMC for permit compliance assessment.

References

American Public Health Association (APHA), American Water Works Association (AWWA) Water Environment Federation (WEF) 1995. *Standard Methods for the Examination of Water and Wastewater*, 19th Edition.

American Sigma 2008. *Sigma 900 MAX Portable Sampler, Sigma Portable Sampler Overview*. http://www.hach.com/hc/view.document.invoker/VendorProductCode=8930/View=HTMLCAT_SIGMA_PORTABLE_SAMPLER_OVERVIEW.

BMP Database Project Team 2002. *Urban Stormwater BMP Performance Monitoring, A Guidance Manual for Meeting the National Stormwater BMP Database Requirements*. Prepared in cooperation with the United States Environmental Protection Agency (USEPA) and the American Society of Civil Engineers (ASCE) by BMP Database Project Team members Geosyntec Consultants, Denver Urban Drainage and Flood Control District, and ASCE Urban Water Resources research Council. EPA Office of Water, EPA-821-B-02-001.

The Boeing Company Santa Susana Field Laboratory 2006. *Results of the Evaluation of the Grab Versus Flow Weighted Samples and Evaluation of Filtered Versus Unfiltered Samples for Radionuclides*. The Boeing Company, October 2006.

Brown, T., Burd, W. Lewis, J. and G. Chang 1995. *Methods and Procedures in Stormwater Data Collection*. In Stormwater NPDES Related Monitoring Needs, Harry Torno (Ed.), Engineering Foundation Conference Proceedings, American Society of Civil Engineers, New York.

Chang, G., Parish, J. and C. Souer 1990. *The First Flush of Runoff and Its Effect on Control Structure Design*. Environmental Resource Management Division, Department of Environmental and Conservation Services, Austin, Texas.

City of Portland NPDES Stormwater Monitoring Program 1996. *Executive Summary Event History Data from Storms Monitored between May 1994 and March 1995*. Prepared by City of Portland and Woodward-Clyde Consultants.

Federal Highway Administration (FHWA) 2001. *Guidance Manual for Monitoring Highway Runoff Water Quality*. United States Department of Transportation FHWA Publication No. FHWA-EP-01-022.

Hager, M.C. 2001. *Evaluating First Flush Runoff*. Stormwater Magazine, September-October 2001. Forester Media.

Han Y., Lau, S.-L., Kayhanian, M. and M. Stenstrom 2006. *Characteristics of Highway Stormwater Runoff*. Water Environment Research 78, 2377.

Herricks, E., Brent R., Milne, I and I. Johnson 1997. *Assessing the Response of Aquatic Organisms to Short-term Exposure to Urban Runoff*. In Effects of Watershed Development and Management of Aquatic Ecosystems, Ed. Larry A. Roesner. American Society of Civil Engineers, Engineering Foundation, New York.

Maestre, A., Pitt R. and D. Williamson 2004. *Nonparametric Statistical Tests Comparing First Flush and Composite Samples from the National Stormwater Quality Database*. In: *Models and Applications to Urban Water Systems*, Volume 12 (edited by W. James). CHI. Guelph, Ontario, pp. 317 – 338.

National Research Council 2008. *Urban Stormwater Management in the United States*. Committee on Reducing Stormwater Discharge Contributions to Water Pollution, Water Science and Technology Board, Division on Earth and Life Studies, National Research Council, National Academies Press, Washington, D.C.

New South Wales Government (Australia) Department of Environment and Climate Change 2008. *Stormwater First Flush Pollution*. <http://www.environment.nsw.au/mao/stormwater.htm>.

Pitt, R., Maestre, A. and R. Morquecho 2004. *The National Stormwater Quality Database (NSQD, version 1.1)*. Department of Civil and Environmental Engineering, University of Alabama, Tuscaloosa, Alabama.

Roa-Espinosa, A. and R. Bannerman 1995. *Monitoring BMP Effectiveness at Industrial Sites*. In Stormwater NPDES Related Monitoring Needs, Harry Torno (Ed.), Engineering Foundation Conference Proceedings, American Society of Civil Engineers, New York.

Scheuler, T. and H. Holland 2000. *Article 9, First Flush of Stormwater Pollutants Investigated in Texas*. Technical Note # 28 from Watershed Protection Techniques. 1(2): 88-89 by TRS. Published in The Practice of Watershed Protection, Center for Watershed Protection.

Southern California Stormwater Monitoring Coalition 2004. *Model Monitoring Program for Municipal Separate Storm Sewer Systems in Southern California*. Technical Report #419. Stormwater Monitoring Coalition Model Monitoring Technical Committee including Regional Water Quality Control Boards, Municipal Permittees, Heal the Bay and Southern California Coastal Water Research Project.

State of California Los Angeles Regional Water Quality Control Board 2007a. *Cease and Desist Order No. R4-2007-0055 Waste Discharge Requirements for The Boeing Company (Santa Susana Field Laboratory) (NPDES No. CA0001309)*. California Regional Water Quality Control Board, Los Angeles Region. Revised: November 1, 2007.

State of California Los Angeles Regional Water Quality Control Board 2007b. *Order No. R4-2007-0056*. California Regional Water Quality Control Board, Los Angeles Region. Revised: November 1, 2007.

State of California Los Angeles Regional Water Quality Control Board 2007c. *Fact Sheet National Pollutant Discharge Elimination System Permit for The Boeing Company (Santa Susana Field Laboratory) (NPDES No. CA0001309)*. California Regional Water Quality Control Board, Los Angeles Region. Revised: October 15, 2007.

State of California Los Angeles Regional Water Quality Control Board 2007d. *Monitoring and Reporting Program No. 6027 for the Boeing Company Santa Susana Field Laboratory (CA0001309)*. California Regional Water Quality Control Board, Los Angeles Region. Revised: November 1, 2007.

Stein, E.D. and V.K. Yoon 2007. *Assessment of Water Quality Concentrations and Loads from Natural Landscapes*. Southern California Coastal Water Research Project, Technical Report 500.

Stenstrom, M.K. and M. Kayhanian 2005. *First Flush Phenomenon Characterization*. Prepared for California Department of Transportation, Division of Environmental Analysis, CTSW-RT-05-73-02.6.

Technology Acceptance Reciprocity Partnership (TARP) 2001 (Updated 2003). *Protocol for Stormwater Best Management Practice Demonstrations*. Endorsed by California, Massachusetts, Maryland, New Jersey, Pennsylvania, and Virginia.

Teledyne ISCO 2004. *Surface Water Monitoring Guide*. Teledyne ISCO, Lincoln, Nebraska.

United States Environmental Protection Agency 2008. *BMP Performance Webcast*. http://www.fedcenter.gov/Bookmarks/index.cfm?id=9049&page_id=1854.

Washington State Department of Ecology 2002. *How to Do Stormwater Sampling: A Guide for Industrial Facilities*. Washington State Department of Ecology Publication #02-10-071.

Additional References:

American Society of Civil Engineers (ASCE), United States Environmental Protection Agency (EPA). *Urban Stormwater BMP Performance Monitoring. A Guidance Manual for Meeting the National Stormwater BMP Database Requirements*. Office of Water. Washington D.C. 2002.

Bailey R. Using Automatic Samplers to comply with stormwater regulations. *Water Engineering and Management*. In: *Water Engineering and Management* (140): 23 – 25. 1993.

Berthouex, P., Brown, L. *Statistics for Environmental Engineers*. Boca Raton : Lewis Publishers. 2002.

Burton, G.A. Jr., and R. Pitt. *Stormwater Effects Handbook: A Tool Box for Watershed Managers, Scientists, and Engineers*. ISBN 0-87371-924-7. CRC Press, Inc., Boca Raton, FL. 2002. 911 pages.

Clark, S.E., C.Y.S. Siu, R.E. Pitt, C.D. Roenning, and D.P. Treese. “Peristaltic pump autosamplers for solids measurement in stormwater runoff.” *Water Environment Research*, in press. 2008.

Leecaster M., Schiff K., Tiefenthaler L. Assessment of efficient sampling designs for urban stormwater monitoring. *Water Research* 36 (2002). Elsevier Science Ltd. 1556 – 1564. 2002.

Maestre, A. and R. Pitt. *The National Stormwater Quality Database, Version 1.1, A Compilation and Analysis of NPDES Stormwater Monitoring Information*. U.S. EPA, Office of Water, Washington, D.C. (final draft report) August 2005.

Pitt, R. *Characterizing and Controlling Urban Runoff through Street and Sewerage Cleaning*. U.S. Environmental Protection Agency, Storm and Combined Sewer Program, Risk Reduction Engineering Laboratory. EPA/600/S2-85/038. PB 85-186500. Cincinnati, Ohio. 467 pgs. June 1985.

Pitt, R. and G. Shawley. *A Demonstration of Non-Point Source Pollution Management on Castro Valley Creek*. Alameda County Flood Control and Water Conservation District and the U.S. Environmental Protection Agency Water Planning Division (Nationwide Urban Runoff Program). Washington, D.C. June 1982.

Pitt, R. and U. Khambhammettu. *Field Verification Tests of the UpFlow™ Filter. Small Business Innovative Research, Phase 2 (SBIR2) Report*. U.S. Environmental Protection Agency, Edison, NJ. 275 pages. March 2006.

Selbig, W.R., and Bannerman, R.T., 2008. *A Comparison of Runoff Quantity and Quality from Two Small Basins Undergoing Implementation of Conventional- and Low-Impact- Development (LID) Strategies: Cross Plains, Wisconsin, Water Years 1999–2005*. U.S. Geological Survey Scientific Investigations Report 2008–5008, 57 pgs. <http://pubs.water.usgs.gov/sir2008-5008>

Appendix A: Los Angeles Regional Rainfall Characteristics Affecting Stormwater Monitoring

This appendix reviews regional rainfall data that affects stormwater monitoring, specifically the time when the rain events start and their durations, and how many events may occur during the nighttime/dusk/dawn hours vs. daylight hours during the rainy season. As noted in the main body of this white paper, a sampling strategy of obtaining a single grab sample at the beginning of the storm was intended to increase the likelihood of including higher concentrations than if the sample was obtained later in the event. However, safety of the personnel obtaining the samples is also critical, so these samples are only obtained during daylight hours. If an event started before daylight, a sample was obtained as soon as possible in the morning, even though that could be several hours after the start of the runoff event. The rain analyses in this appendix were therefore conducted to determine the likely numbers of events that may occur during light and dark conditions.

Fifty years (1949 through 1999) of LAX rainfall records were examined to identify the likely starting times of various rainfall depths. Hourly rainfall values from the National Weather Service were obtained from EarthInfo CD Roms and processed in WinSLAMM to create separate rain events. Each rain event was defined as having at least 0.01 inches of rain in one hour, with an interevent dry period of at least 6 hours. WinSLAMM prepared a statistical summary of the monthly rainfall data, with some of the data shown in Table A-1.

Table A-1. Monthly Rainfall Variations for LAX from 1949 through 1999

Month	Average rainfall depth, per month, inches	Monthly rainfall depth coefficient of variation	Average number of rain events per month (>0.01 inch each)	Monthly rainfall count coefficient of variation
January	3.0	0.96	5.4	0.63
February	2.8	1.1	4.8	0.67
March	1.8	0.93	4.9	0.71
April	0.78	1.2	2.8	0.94
May	0.18	2.8	1.2	1.5
June	0.05	2.8	0.4	1.6
July	0.02	2.7	0.4	1.8

August	0.09	4.1	0.4	2.0
September	0.21	2.3	1.1	1.3
October	0.28	1.7	1.4	1.0
November	1.5	1.1	3.1	0.67
December	1.7	0.92	4.0	0.64
Total	12.5	0.46 (calc. from 50 years of annual totals)	29.9	0.30 (calc. from 50 years of annual totals)

Although not shown on the above table, most of these rains had durations of several hours (3-5 hrs), although the very smallest rains had durations of only 1 hour and rains larger than about one inch had durations of about one day (24 hrs).

As shown, November through March have most of the rains in this area (10.9 inches out of 12.5 inches and 22 out of 30 events on the average), based on these 50 years of LAX data. However, the variation can be quite large, as reflected in the large coefficient of variation values (COV). The COVs are much larger for months having little rainfall. The annual total rain depths and numbers of events based on the 50 year totals have much less variation than the individual monthly values.

Daylight hours were obtained from SkyTools (version 2, CapellaSoft) for Simi Valley for the mid months for this rainy season, as shown in Table A-2. During this five month period, the typical daylight hours are from about 0700 to 1700 PST.

Table A-2. Daylight Hours at Simi Valley (standard time at the 15th of each month)

	Nov	Dec	Jan	Feb	Mar
sunrise	06:29	06:54	06:22	06:38	06:03
sunset	16:50	16:46	16:12	17:39	18:03

The start times for the more than 1500 rains during this 50 year period are shown in Figure A-1. This figure does not show much of a pattern because of the overlapping data points and large scatter.

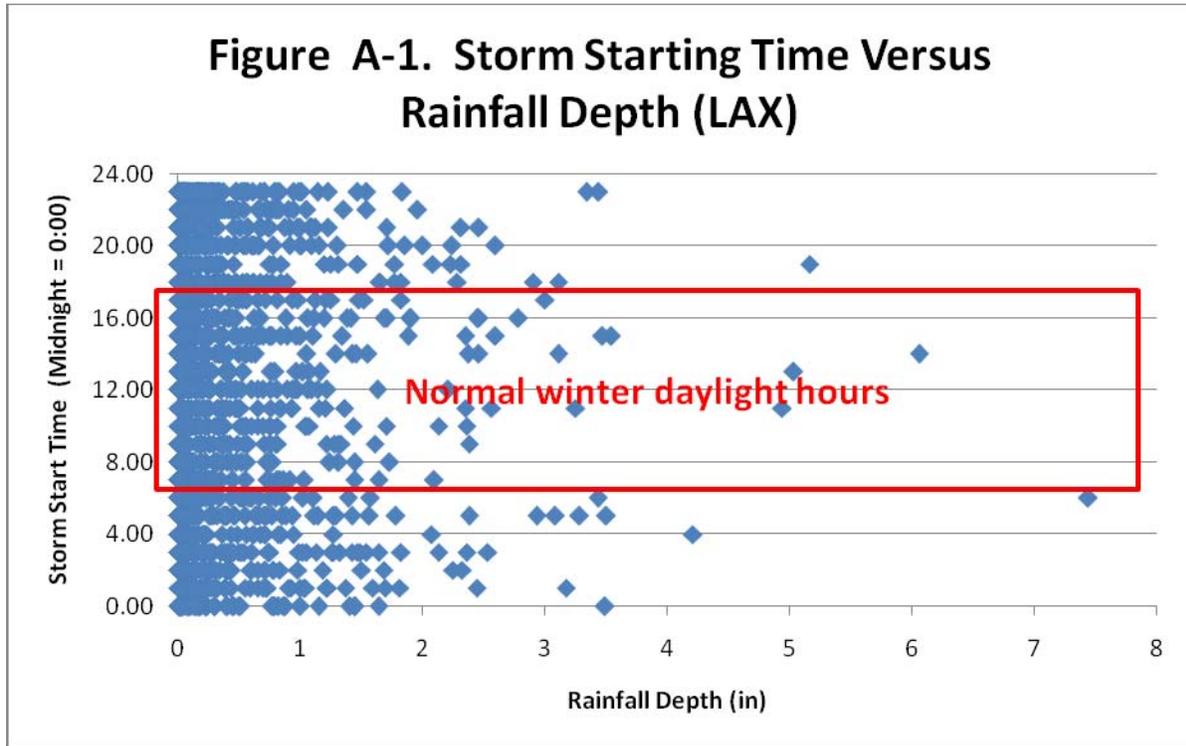


Figure A-1. Plot of starting time vs. rain depth

Figure A-2 is a histogram showing the average percentage of the total rains that started in each hour of the day. This figure does show a large difference in the hour to hour rain starts, with the early morning and late evening hours having many more events than the midday hours.

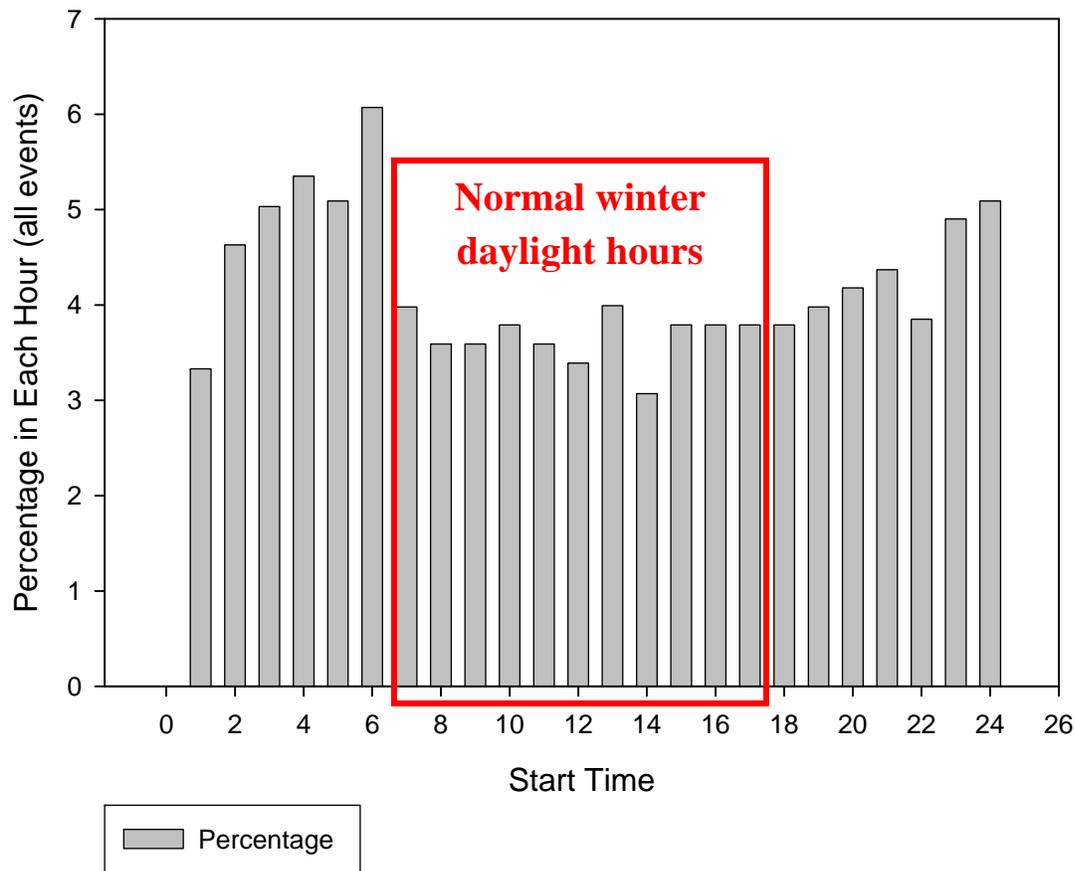


Figure A-2. Percentage of rains starting in each hour.

The data were therefore further processed to determine the percentage of the rains starting in each hour for different rain classes, as summarized in Table A-3. This table also shows the total number of rains in each depth category, along with the numbers of rains occurring in daylight and nighttime/dusk hours. For all rain depth conditions, the daylight hours (about 37%) have many fewer events than the nighttime/dusk hours (about 63%).

The numbers of events available to sample during daylight hours is relatively low, ranging from about 8 events per year for all events, to only about 3 to 5 events per year for rains that likely generate runoff (estimated to be >0.25 to >0.5 inch rains, depending on antecedent moisture conditions, rain intensity, and other factors). Obviously, year to year variations can be quite large.

It is expected that nighttime/dusk sampling in addition to the daytime sampling would increase the average number of events per year that can be sampled to about 10 events, at least. Therefore, in order to obtain representative data from as many storm events per year as possible needed to characterize site runoff conditions, safe sampling needs to occur at all hours.

Table A-3. Percentage of All Events (by count) Starting at Different Hours of the Day

start hour	all events (>0.01")	events >0.25"	events >0.50"	events >0.75"	events >1.0"	events >2.5"
0	3.3	3.5	4.1	5.2	3.8	3.7
1	4.6	4.8	4.9	4.8	6.0	3.7
2	5.0	4.4	3.6	3.3	2.7	0.0
3	5.4	6.0	5.9	6.3	8.2	3.7
4	5.1	2.4	3.1	2.2	1.6	3.7
5	6.1	5.6	5.9	5.9	6.6	14.8
6	4.0	3.2	3.9	3.3	3.3	7.4
7	3.6	3.4	2.8	3.3	2.2	0.0
8	3.6	2.7	2.3	1.9	2.2	0.0
9	3.8	4.0	3.3	3.7	3.3	0.0
10	3.6	3.9	4.4	3.7	3.8	0.0
11	3.4	4.5	4.1	4.8	4.4	11.1
12	4.0	4.0	4.1	4.5	3.3	0.0
13	3.1	2.4	2.1	2.6	2.2	3.7
14	3.8	3.5	3.6	3.3	4.9	7.4
15	3.8	5.0	5.7	5.2	4.4	11.1
16	3.8	4.0	4.1	4.8	6.6	3.7
17	3.8	3.4	4.4	4.1	4.4	3.7

18	4.0	5.0	4.4	3.7	3.8	7.4
19	4.2	4.2	3.6	5.2	5.5	3.7
20	4.4	5.2	5.4	4.5	4.9	3.7
21	3.9	3.7	4.9	4.8	4.9	0.0
22	4.9	5.5	4.1	3.7	2.2	0.0
23	5.1	5.5	5.4	4.8	4.4	7.4
average total # of events from Nov to Mar per year	22	12.4	7.8	5.4	3.6	1
% (#) during daylight hours	36 (8)	38 (4.7)	37 (2.8)	38 (2.0)	37 (1.4)	37 (0.3)
% (#) during nighttime/dusk hours	63 (14)	62 (7.7)	63 (4.9)	62 (3.3)	63 (2.3)	63 (0.6)

Appendix B: Peak Concentrations vs. Representative Concentrations during the Entire Runoff Event

Water quality criteria associated with different beneficial uses have different “averaging” times. These averaging times are usually one hour average values for “maximum” concentrations and 4-days for continuous concentrations to protect freshwater aquatic life. As would be expected using standard toxicological relationships, the short duration criteria values are larger than the long duration criteria values. Human health standards associated with carcinogens are usually based on life-time exposures calculated using an assumed consumption of fish and shellfish, and consumption of water. The standards are long-duration exposures associated with the calculated uptake of the pollutants into the organisms and the assumed amount of fish consumed, or the contamination of a drinking water source.

Most discharge goals associated with stormwater are related to site-specific Basin Plans that recognize water quality standards not being met in a receiving water and the water quality goals for the beneficial uses. Necessary reductions in discharges to meet the water quality standards are used to calculate the Total Maximum **Daily** Loads (TMDLs) for different discharge categories (industrial and municipal point source discharges and stormwater nonpoint source discharges). As indicated, these are daily loads that are usually calculated for a critical low flow period in the receiving water (such as the 7Q10, the lowest 7 day average flow period that reoccurs every decade). Continuous point sources can be directly related to these discharge limits, but it is more difficult to relate intermittent stormwater discharges to these goals, especially as stormwater is not likely being discharged during these low flow periods. The “averaging” time for stormwater discharge limits specified in the TMDL reports are therefore at least one day to a week, and relate to expected mass discharges of pollutants. The averaging times for monitored stormwater associated with TMDL discharge limits are therefore most closely related to the complete event mass discharges and not peak concentrations.

As shown on the following table having the Santa Susana stormwater discharge limits, most of the constituent limits are derived from the basin plans or TMDLs. In fact, the numeric values are specifically shown as daily maximums. Daily maximums, derived from TMDLs and basin plans, are associated with 24-hour duration monitoring (averaging) periods. These are not the maximum concentrations that occur within any day. Others (Cd, Cu, Hg, Pb, and TCDD) are associated with the California Toxics Rule (CTR). In the CTR, criteria are established for “continuous” or “acute” exposures. In the April 1999 compilation report (*National Recommended Water Quality Criteria-Correction*, US EPA, EPA 822-Z-99-001), CMC refers to the “Criterion Maximum Concentration” with an exposure period of one hour (generally corresponding to the earlier “acute” criterion), and CCC refers to the “Criterion Continuous Concentration” with an averaging period of 4 days (generally corresponding to the earlier “chronic” criterion). As stated in 40 C.F.R. § 131.38, CCC is the water quality criteria to protect against chronic effects in aquatic life and is the highest in-stream concentration of a priority toxic pollutant consisting of a 4-day average not to be exceeded more than once every three years on

the average. The CMC equals the highest concentration of a pollutant to which aquatic life can be exposed for a short period of time without deleterious effects. As noted in the footnotes in the CFR table, the acute values for the pesticides lindane, dieldrin, and endrin are taken as instantaneous values, based on the 1980 guidelines. However, these pesticides are not included in the Santa Susana stormwater discharge permit. Therefore, most of the site stormwater permit limits were derived for daily (24 hr) averaging periods, while Cd, Cu, Hg, Pb, and TCDD limits may have been derived for one hour limits, although the permit limits specify daily maximum periods for all constituents. However, the permit notes single grab samples to be taken during the first part of the runoff event, during a period of assumed peak concentrations.

Santa Susana Stormwater Discharge Permit Limits

Constituents	Units	Discharge Limitations		Rationale
		Monthly Average	Daily Maximum	
pH	pH Units	----	6.5-8.5	Basin Plan
Oil and grease	mg/L	----	15	BPJ
Chloride	mg/L	----	150	Basin Plan
Sulfate	mg/L	----	250 ²³	Basin Plan
Sulfate	mg/L	----	300 ²⁴	Basin Plan
Boron ¹	mg/L	----	1.0	Basin Plan
Fluoride	mg/L	----	1.6	Basin Plan
Nitrate + Nitrate-N	mg/L	----	10.0 ²⁵	Basin Plan
Nitrate + Nitrate-N	mg/L	----	8 ²³	Basin Plan
Total dissolved solids	mg/L	----	850 ²⁶	Basin Plan
Total dissolved solids	mg/L	----	950 ²⁶	Basin Plan
Ammonia-N ^(consist only)	mg/L	----	10.1 ²⁷	LA River Nutrients TMDL
Nitrate-N ^(consist only)	mg/L	----	8.0	LA River Nutrients TMDL
Nitrite-N ^(consist only)	mg/L	----	1.0	LA River Nutrients TMDL
Selenium ^(consist only)	µg/L	----	5 ²⁸	LA River Metals TMDL
Zinc ^(consist only)	µg/L	----	159 ²⁹	LA River Metals TMDL
Antimony	µg/L	----	6.0	Basin Plan/Title 22
Cadmium	µg/L	----	4.0/3.1 ³⁰	CTR/LA River Metals TMDL
Copper	µg/L	----	14.0	CTR
Mercury	µg/L	----	0.13	Calleguas Creek Metals TMDL (CTR)
Nickel	µg/L	----	100	Calleguas Creek Metals TMDL/Basin Plan (Title 22)
Thallium	µg/L	----	2.0	Basin Plan
Lead	µg/L	----	5.2	CTR
TCDD	µg/L	----	2.8E-08	CTR
Perchlorate	µg/L	----	6.0	BPJ/ DHS Notification Level
Radioactivity Gross Alpha	pCi/L	----	15	Basin Plan/Title 22
Gross Beta	pCi/L	----	50	Basin Plan/Title 22
Combined Radium-226 & Radium-228	pCi/L	----	5	Basin Plan/Title 22
Tritium	pCi/L	----	20,000	Basin Plan/Title 22

Final California Toxic Rule Criteria (August 5, 1997 Federal Register, 40 CFR Part 131, available at: <http://www.epa.gov/fedrgstr/EPA-WATER/1997/August/Day-05/w20173.pdf>)

Compound	Proposed freshwater		NTR freshwater	
	CMC	CCC	CMC	CCC
Arsenic	1.2 340	1.2 150	1.3 360	1.3 190
Cadmium	1.2,4 3	1.2,4 2.2	1.5 3.7	1.5 1.0
Chromium (VI)	1.2 16	1.2 11	1.3 15	1.3 10
Copper	1.2,4 13	1.2,4 9.0	1.5 17	1.5 11
Nickel	1.2,4 470	1.2,4 52	1.5 1400	1.5 160
Zinc	1.2,4 120	1.2,4 120	1.5 110	1.5 100
Pentachloro-phenol	2.6 19	2.6 15	6 20	6 13
Lindane (gamma-BHC)	≥ 0.95	7 2	0.08
Dieldrin	≥ 0.24	≥ 0.056	7 2.5	0.0019
Endrin	≥ 0.086	≥ 0.036	7 0.18	0.0023
Mercury	1.2,3 1.4	1.2,3 0.77	1.3 2.1	0.012
PCBs	8,9 0.014	8,10 0.014
Mercury	1.3 1.8	1.3 0.94	1.3 1.8	0.025
PCBs	8,9 0.03	8,10 0.03

¹ These freshwater and saltwater criteria for metals are expressed in terms of the dissolved fraction of the metal in the water column, not the total recoverable fraction. Criterion values were calculated by using EPA's CWA 304(a) criteria guidance values (described in the total recoverable fraction) and then applying conversion factors as in the NTR, as amended, (60 FR 22228, May 4, 1995 and 40 CFR part 131).

² This criterion has been recalculated pursuant to 1995 Updates: *Water Quality Criteria Documents for the Protection of Aquatic Life in Ambient Water* (EPA-820-B-96-001, September 1996). See also the Great Lakes Water Quality Initiative (40 CFR Parts 9, 122, 123, 131, and 132; Final Water Quality Guidance for the Great Lakes System, Final Rule: 60 FR 15366, March 23, 1995) and *Great Lakes Water Quality Initiative Criteria Documents for the Protection of Aquatic Life in Ambient Water* (EPA-820-B-95-004, March 1995).

³ Criteria for these metals are expressed as a function of the water-effect ratio (WER) as defined in 40 CFR 131.36(c).

⁴ These freshwater aquatic life criteria for metals are expressed as a function of total hardness (mg/l as CaCO₃) in the water body. The equations are provided in the proposed rule at 40 CFR 131.38(b)(2). Values displayed above and in the proposed rule matrix correspond to a total hardness of 100 mg/l as CaCO₃.

⁵ Freshwater aquatic life criteria for these metals are expressed as a function of total hardness (mg/l as CaCO₃), and as a function of the pollutant's water-effect ratio, WER, as defined in 40 CFR 131.36(c). The equations are provided in the NTR, as amended, and values above and in the rule matrix correspond to a total hardness of 100 mg/l as CaCO₃ and a water-effect ratio of 1.0.

⁶ These freshwater aquatic life criteria for pentachlorophenol are expressed as a function of pH, and are calculated as follows: (Values displayed above in the matrix correspond to a pH of 7.8.) CMC=exp(1.005(pH)-4.830). CCC=exp(1.005(pH)-5.290).

⁷ These aquatic life criteria for these pollutants were issued in 1980 utilizing the 1980 Guidelines for criteria development. The acute values shown are final acute values (FAV) which by the 1980 Guidelines are instantaneous values.

⁸ The CAS numbers for the PCB compounds are 53469219, 11097691, 11104282, 11141165, 12672296, 11096825, and 12674112, respectively.

⁹ This proposed criterion is the sum of all aroclors.

¹⁰ This criterion was listed for each aroclor in the matrix at 40 CFR 131.36(b)(1).

Because stormwater discharge limits and resulting “averaging” periods can be ambiguous (ranging from 1 hour for some acute toxicants to daily totals for TMDL-derived limits), it may be best to have a flexible monitoring strategy, in the absence of specific guidance from the regulatory agency. As an example, the initial 1987 Federal Register sampling guidance for stormwater permits under the NPDES program recommended a two part monitoring effort: one sample during the first 30 minutes of the runoff event and a composite sample during the complete event (but strangely, only for the first 3 hours of runoff duration). This monitoring guidance could be modified by delegated authorities. In fact, most states did not require the initial “first flush” sample in the required monitoring programs. The initial grab sample was used for the analysis of the “first flush effect,” which assumes that more of the pollutants are discharged during the first period of runoff than during later periods. The following is a summary of calculations made from examining events that included separate samples collected during both the first 30 minutes and for the first 3 hours of the event (the composite sample), as obtained in the National Stormwater Quality Database (NSQD) (Maestre and Pitt 2005).

First 30 Minute vs. First 3 Hour Stormwater Characteristics from the NSQD

A total of 417 storm events having paired first flush and composite samples were available from the NPDES MS4 database. The majority of the events were located in North Carolina (76.2%), but some events were also from Alabama (3.1%), Kentucky (13.9%) and Kansas (6.7%). All of the events corresponded to end-of-pipe samples in separate storm drainage systems. Most of the data were from residential, commercial, and industrial land use sites, with fewer data from open space, institutional, and freeway land uses.

About 83% of the possible paired cases were evaluated. The remaining cases could not be evaluated because the data set did not have enough paired data or they did not fit the requirements of the Fligner-Policello, the modified T-test, or the Mann-Whitney statistical tests that were used for these analyses. Table B-1 shows the results of the analysis for commercial, industrial and institutional land use areas.

Table B-1. Significant First Flushes Ratios (first flush to composite median concentration)

Parameter	Commercial				Industrial				Institutional			
	n	sc	R	ratio	n	sc	R	ratio	n	sc	R	ratio
Turbidity, NTU	11	11	=	1.32			X				X	
pH, S.U.	17	17	=	1.03	16	16	=	1.00			X	
COD, mg/L	91	91	≠	2.29	84	84	≠	1.43	18	18	≠	2.73
TSS, mg/L	90	90	≠	1.85	83	83	=	0.97	18	18	≠	2.12
BOD ₅ , mg/L	83	83	≠	1.77	80	80	≠	1.58	18	18	≠	1.67
TDS, mg/L	82	82	≠	1.83	82	81	≠	1.32	18	18	≠	2.66
O&G, mg/L	10	10	≠	1.54			X				X	
Fecal Coliform, col/100mL	12	12	=	0.87			X				X	
Fecal Streptococcus, col/100 mL	12	11	=	1.05			X				X	
Ammonia, mg/L	70	52	≠	2.11	40	33	=	1.08	18	16	≠	1.66
NO ₂ + NO ₃ , mg/L	84	82	≠	1.73	72	71	≠	1.31	18	18	≠	1.70
N Total, mg/L	19	19	=	1.35	19	16	=	1.79			X	
TKN, mg/L	93	86	≠	1.71	77	76	≠	1.35			X	
P Total, mg/L	89	77	≠	1.44	84	71	=	1.42	17	17	=	1.24
P Dissolved, mg/L	91	69	=	1.23	77	50	=	1.04	18	14	=	1.05
Ortho-P, mg/L			X		6	6	=	1.55			X	
Cadmium Total, µg/L	74	48	≠	2.15	80	41	=	1.00			X	

Chromium Total, µg/L	47	22	≠	1.67	54	25	=	1.36			X	
Copper Total, µg/L	92	82	≠	1.62	84	76	≠	1.24	18	7	=	0.94
Lead Total, µg/L	89	83	≠	1.65	84	71	≠	1.41	18	13	≠	2.28
Nickel, µg/L	47	23	≠	2.40	51	22	=	1.00			X	
Zinc, µg/L	90	90	≠	1.93	83	83	≠	1.54	18	18	≠	2.48
Turbidity, NTU			X		12	12	=	1.24	26	26	=	1.26
pH, S.U.			X		26	26	=	1.01	63	63	=	1.01
COD, mg/L	28	28	=	0.67	140	140	≠	1.63	363	363	≠	1.71
TSS, mg/L	32	32	=	0.95	144	144	≠	1.84	372	372	≠	1.60
BOD5, mg/L	28	28	=	1.07	133	133	≠	1.67	344	344	≠	1.67
TDS, mg/L	31	30	=	1.07	137	133	≠	1.52	354	342	≠	1.55
O&G, mg/L			X				X		18	14	≠	1.60
Fecal Coliform, col/100mL			X		10	9	=	0.98	22	21	=	1.21
Fecal Streptococcus, col/100 mL			X		11	8	=	1.30	26	22	=	1.11
Ammonia, mg/L			X		119	86	≠	1.36	269	190	≠	1.54
NO ₂ + NO ₃ , mg/L	30	21	=	0.96	121	118	≠	1.66	324	310	≠	1.50
N Total, mg/L	6	6	=	1.53	31	30	=	0.88	77	73	=	1.22
TKN, mg/L	32	14	=	1.28	131	123	≠	1.65	335	301	≠	1.60
P Total, mg/L	32	20	=	1.05	140	128	≠	1.46	363	313	≠	1.45
P Dissolved, mg/L	32	14	=	0.69	130	105	≠	1.24	350	254	=	1.07
Ortho-P, mg/L			X		14	14	=	0.95	22	22	=	1.30
Cadmium Total, µg/L	30	15	=	1.30	123	33	≠	2.00	325	139	≠	1.62
Chromium Total, µg/L	16	4	=	1.70	86	31	=	1.24	218	82	≠	1.47
Copper Total, µg/L	30	22	=	0.78	144	108	≠	1.33	368	295	≠	1.33
Lead Total, µg/L	31	16	=	0.90	140	93	≠	1.48	364	278	≠	1.50
Nickel, µg/L			X		83	18	=	1.20	213	64	≠	1.50
Zinc, µg/L	21	21	=	1.25	136	136	≠	1.58	350	350	≠	1.59

Note: n = number of total possible events. sc = number of selected events with detected values. R = result. Not enough data (X); not enough evidence to conclude that median values are different (=); median values are different (≠).

The “≠” sign indicates that the medians of the first flush and the composite data set are different. The “=” sign indicates that there is not enough information to reject the null hypothesis at the desired level of confidence (at least at the 95% level). Events without enough data are represented with an “X”.

Also, shown on this table are the ratios of the medians of the first flush to the composite data for each constituent and land use combination. Generally, a statistically significant first flush is associated with a median concentration ratio of about 1.4, or greater (the exceptions are where the number of samples in a specific category is much smaller). The largest ratios are about 2.5, indicating that for these conditions, the first flush sample concentrations are about 2.5 times greater than the composite sample concentrations. More of the larger ratios are found for the commercial and institutional land use categories, areas where larger paved areas are likely to be found. The smallest ratios are associated with the residential, industrial, and open spaces land uses, locations where there may be larger areas of unpaved surfaces.

Results indicate that for 55% of the evaluated cases, the median of the first flush data set were greater than the composite sample set. In the remaining 45% of the cases, both medians were likely the same (not enough data to prove that they are different), or the concentrations were possibly greater later in the events.

Approximately 70% of the constituents in the commercial land use category had elevated first flush concentrations, about 60% of the constituents in the residential, institutional and the mixed (mostly commercial and residential) land use categories had elevated first flushes, and about 45% of the constituents in the industrial land use category had elevated first flushes. In contrast, no constituents were found to have elevated first flushes in the open space category (all located in North Carolina).

COD, BOD₅, TDS, TKN and Zn all had first flushes in all areas (except for the open space category). In contrast, turbidity, pH, fecal coliform, fecal streptococcus, total N, dissolved and ortho-P never showed a statistically significant first flush in any category. The different findings for TKN and total nitrogen imply that there may be other factors involved in the identification of first flushes besides land use.

It is expected that peak concentrations generally occur during periods of peak flows (and highest rain energy). On relatively small paved areas, however, it is likely that there will always be a short initial period of relatively high concentrations associated with washing off of the most available material (Pitt 1987). This peak period of high concentrations may be overwhelmed by periods of high rain intensity that may occur later in the event. In addition, in more complex drainage areas, the routing of these short periods of peak concentrations may blend with larger flows and may not be noticeable. A first flush in a separate storm drainage system is therefore most likely to be seen if a rain occurs at relatively constant intensities on a paved area having a

simple drainage system. If the peak flow (and highest rain energy) occurs later in the event, then there likely will not be a noticeable first flush. However, if the rain intensity peak occurs at the beginning of the event, then the effect is exaggerated. Groups of constituents also showed different behaviors for different land uses. All the heavy metals evaluated showed higher concentrations at the beginning of the events in the commercial land use category. Similarly, all of the nutrients showed higher initial concentrations in residential land use areas, except for total nitrogen and ortho-phosphorus. None of the land uses showed a higher population of bacteria during the beginning of the event. Conventional constituents showed elevated first-flush concentrations in commercial, residential and institutional land uses.

The data available for these analyses were mostly from the southeast and it is not known how transferable these findings would be for other areas of the country. However, the general findings are cause of interest and indicate that elevated concentrations are not always at the beginning of runoff events, especially in open space areas.

Appendix C. Potential Problems Associated with Manual and Automatic Water Sampling and how to Overcome Them

Problems Encountered during NPDES Stormwater Monitoring, as Reported in the NSQD and other Sources

When compiling data for the National Stormwater Quality Database (NSQD), summaries of reported sampling problems were also prepared by Maestre and Pitt (2005). These problems associated with compliance monitoring by various communities as part of their stormwater permits are described in the following discussion and indicate some of the short-comings associated with different sampling schemes.

The over-specificity of sampling requirements in the discharge permits that do not consider local conditions was an obvious problem. About 58% of the communities submitting data for the NSQD described problems they had during the monitoring process and were summarized in the annual monitoring reports. One of the basic sampling requirements was to collect three samples every year for each of the land use stations. These samples were to be collected at least one month apart during rains having at least 0.1 inch rains, and with at least 72 hours from the previous 0.1-inch storm event. It was also required (when feasible), that the variance in the duration of the event and the total rainfall not exceeded the median rainfall for the area. About 47% of the communities reported problems meeting these requirements. In many areas of the country, it was difficult to have three storm events per year meeting these requirements.

Errors in the siting and installation of the monitoring equipment cause sampling problems. The second most frequent problem, reported by 26% of the communities, concerned backwater (or tidal) influences during sampling, when the outfall became submerged during the event. In other cases, it was observed that there was flow under the pipe (flowing outside of the pipe, in the backfill material, likely groundwater), or sometimes there was not flow at all at the outfall during obvious rains. More care needs to be taken when sitting sampling equipment and careful surveys of infrastructure are needed in developed areas to ensure that the sampling locations are suitable and representative.

Equipment malfunctions are common and monitoring stations require frequent maintenance. About 12% of the communities described errors related to malfunctions of the sampling equipment. When reported, the equipment failures described were due to incompatibility between the software and the equipment, clogging of the rain gauges, and obstruction in the sampling or bubbler lines. Memory losses in the equipment recording data were also periodically reported. Other reported problems were associated with lightning, false starts of the automatic sampler before the runoff started, and operator error due to misinterpretation of the equipment configuration manual.

Setting up samplers to represent a wide range of rain depths is challenging and sufficient rain gauges on site are needed. Capturing runoff events within the acceptable range of rain

depth was difficult for some monitoring agencies. Rain depth cannot be accurately predicted in many areas of the country. Also, if using rain gauge data from a location distant from the monitoring location, the reported rain depth may not have been representative of the rain conditions that occurred at the site. The rain gauges need to be placed close to the monitored watersheds. This was likely one of the reasons why the runoff depths periodically exceeded the reported rain depths. Rain in urban areas can vary greatly over small distances. The ASCE/EPA (2002) recommended that rainfall gauges be located as close as possible to the monitoring station. Another factor that needs to be considered is the size of the watershed. Large watersheds cannot be represented with a single rain gauge at the monitoring station; in those cases the use of monitoring networks will be a better approach. Large watersheds are more difficult to represent with a single rain depth value. Setting up automatic samplers to represent a wide range of event sizes is difficult with standard commercial equipment.

Accurate flow monitoring equipment needed. Many of the monitoring stations lacked flow monitoring instrumentation, or did not properly evaluate the flow data. Accurate flow monitoring can be difficult, but it greatly adds to the value of the expensive water quality data. In addition, recent work by the USGS (described below) has found that accurate calibration of stormwater flow monitoring equipment is needed. They developed automatic dye injection systems to calibrate flows during actual rains that work well for large sites.

Accurately rated flumes require minimal calibration, beyond water level checks. If possible, flow rates should be verified under a range of conditions. However, if the flume is correctly installed, the water level measurements should be the only measurements needing adjustment. However, many stormwater monitoring projects now use area-velocity sensors placed in the pipes. Calibration of these sensors can be accomplished during actual runoff periods using a dye tracer and laboratory fluorometer. Selbig and Bannerman (2008) describe this procedure and Figure C-1 illustrates one of their recent calibration efforts. In this method, a known concentration of rhodamine dye was continuously injected at a constant rate sufficiently upstream from the monitoring station to allow for complete mixing during stormwater-runoff events. In some of their urban monitoring stations, the rhodamine dye pump is automatically started when the water rises in the pipe, or after a set amount of rainfall. A stock solution of the partially diluted dye is kept well mixed until the injection starts, and during the injection period. Samples of the dye mixed in the runoff were obtained with a dedicated automated sampler. Discrete samples are collected at equal time increments during both rising and falling water levels. The samples are then taken to the laboratory where the dilution factors of the dye are determined. The dilution factor corresponds to the runoff flow rate.

They reported that velocity sensors using acoustic or electromagnetic detection consistently underestimated actual values. The errors shown in Figure C-1 are not unusual and highlight the importance of *in-situ* calibrating of area-velocity instrumentation.

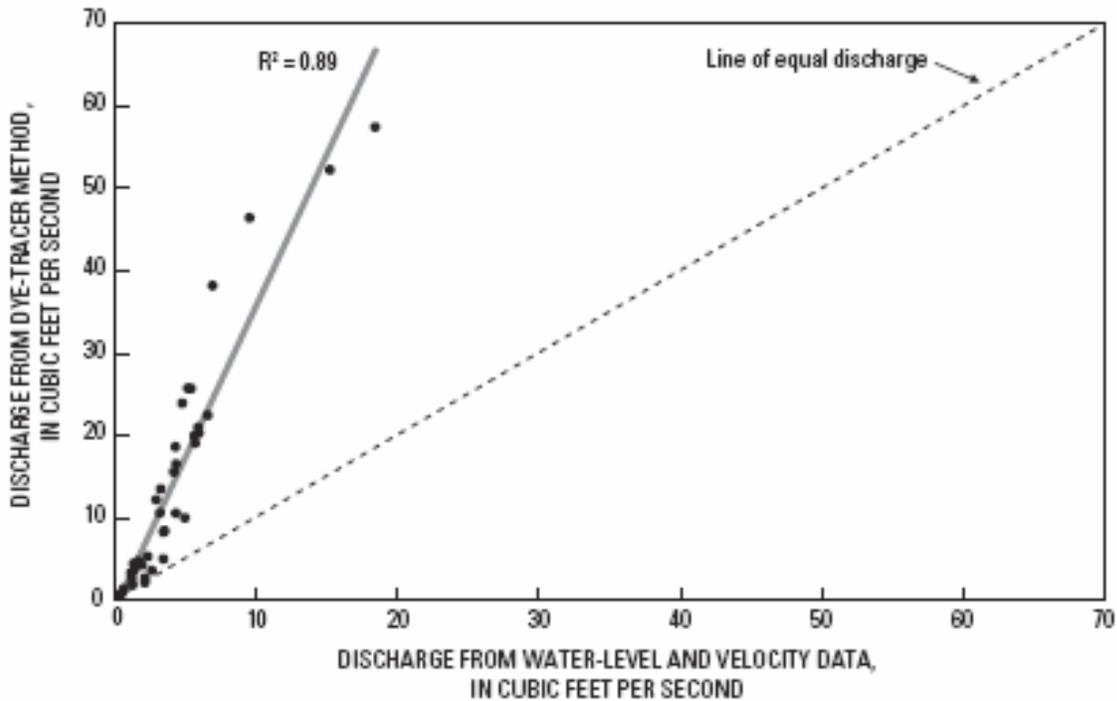


Figure C-1. Relation between actual discharge determined using a rhodamine dye tracer and measured discharge, computed by water-level and velocity data, during free-flow conditions (Selbig and Bannerman 2008).

Short periods of monitoring are not representative of the complete runoff event. The three hour monitoring period that most stormwater permittees used for the “whole” event monitoring likely resulted in some bias in the reported water quality data. This limit was likely used to minimize the length of time personnel needed to be at a monitoring location during manual sampling activities. Also, it is unlikely that manual samplers were able to initiate sampling near the beginning of the events, unless they were deployed in anticipation of an event later in the day. A more cost-effective and reliable option would be to use automatic monitoring equipment located at the monitoring locations and sampling equipment placed on standby in anticipation of a monitored event. A later discussion presents some automatic sampler setup options.

Frequent non-detectable analytical results cause problems in statistical analyses of the data. As the level of non-detected observations increase, the mean, median and standard deviation values are larger than if the censored observations are detected. The opposite behavior is expected for the coefficient of variation. Different laboratories report different detection limits for the same constituents. In the NSQD, open space has the largest number of non-detected observations among the represented land uses. The largest percentages of detected observations were observed in freeways and industrial land uses. Estimating or replacing by half of the detection limit for levels of censoring smaller than 5% does not have a significant effect on the mean, standard deviation and coefficient of variation values. However, replacing the censored

observations by half of the detection limit is not recommended for levels of censoring larger than 15% as the errors in the means and variations can be very large, greatly hindering accurate statistical analyses of the data. Median values are not affected until the non-detected levels exceed 50%, obviously. Further details of problems associated with excessive detection limits in stormwater are discussed by Maestre and Pitt (2005) and substitution options are succinctly described by Berthouex and Brown (2001).

Probability plots of the available stormwater quality data from Santa Susana outfalls 008 and 009 were prepared to estimate the data variability for determining experimental design features (recommended detection limits so less than about 10% of the samples would have non-detected observations and to estimate the number of samples that may be needed for useful results, considering both power and confidence). There were about 20 events at outfall 008 and about 30 events at outfall 009 that were monitored during the 40 months from October 2004 to February 2008 (normally a suitable number of data to determine variations). However, not all events had samples analyzed for all constituents and not all samples had detected observations. Because of the relatively few runoff events that occur at the monitored outfalls (<1 per month), there are few data available for many of the constituents, even with more than three years of monitoring every runoff opportunity. Coupled with the relatively high variability of the stormwater (not unusual), these results therefore need to be carefully interpreted. Table C-1 summarizes the frequency of detected results at these outfalls, the reported detection limits, and the recommended detection limits, along with an indication of the suitability of the detection limits being used to obtain at least a 10% level of non-detected values over a lengthy monitoring period. It is unlikely that some of the permitted constituents can be analyzed at a level to produce the suggested 90% detection level. These constituents that will likely experience excessive detection limit problems are high-lighted in yellow. All of the reported detection limits used by the laboratories for the on-going monitoring program are lower than the permit limits, the most critical objective for the detection limits.

Table C-1. Detection Frequency and Detection Limits of Stormwater Constituents at Outfalls 008 and 009

Constituent	Outfall	Detected (% of total samples)	Detection Limits Reported	Recommended Detection Limit for about 10% Level of Non-detects	Category for detection limit in order to have at least 90% detections (all detection limits that have been used are smaller than the permit levels)
Antimony, total	008	37%	0.05 to 2.5 µg/L	0.1 µg/L	OK, if best used

Antimony, total	009	68%	0.05 to 2.5 µg/L	0.1 µg/L	OK, if best used
Antimony, filtered	008	100%	0.05 to 2.5 µg/L	0.2 µg/L	OK, if best used
Antimony, filtered	009	100%	0.05 to 2.5 µg/L	0.5 µg/L	OK, if best used
Boron	008	100%	0.007 to 0.02 mg/L	0.04 mg/L	OK
Boron	009	100%	0.007 to 0.02 mg/L	0.04 mg/L	OK
Cadmium, total	008	74%	0.015 to 0.12 µg/L	0.001 µg/L	use best available
Cadmium, total	009	65%	0.015 to 0.12 µg/L	0.001 µg/L	use best available
Chloride	008	100%	0.15 to 0.75 mg/L	2.5 mg/L	OK
Chloride	009	100%	0.15 to 0.75 mg/L	2.5 mg/L	OK
Copper, total	008	95%	0.25 to 0.75 µg/L	2.5 µg/L	OK
Copper, total	009	100%	0.25 to 0.75 µg/L	1.5 µg/L	OK
Copper, filtered	008	100%	0.25 to 0.75 µg/L	1.5 µg/L	OK
Copper, filtered	009	100%	0.25 to 0.75 µg/L	2 µg/L	OK
Gross alpha radioactivity	008	100%	0.6 to 2 pCi/L	<<1 pCi/L	use best available
Gross alpha radioactivity	009	100%	0.6 to 2 pCi/L	10 pCi/L	OK
Gross beta radioactivity	008	100%	0.8 to 2 pCi/L	2 pCi/L	OK
Gross beta	009	100%	0.8 to 2	1 pCi/L	OK

radioactivity			pCi/L		
Lead, total	008	100%	0.04 to 0.3 µg/L	0.4 µg/L	OK
Lead, total	009	87%	0.04 to 0.3 µg/L	0.1 µg/L	OK, if best used
Lead, filtered	008	33%	0.04 to 0.3 µg/L	n/a µg/L	n/a
Lead, filtered	009	89%	0.04 to 0.3 µg/L	0.02 µg/L	OK, if best used
Mercury, total	008	37%	0.05 to 0.063 µg/L	0.02 µg/L	OK
Mercury, total	009	32%	0.05 to 0.063 µg/L	0.02 µg/L	OK
Nickel, total	008	100%	2 µg/L	3 µg/L	OK
Nickel, total	009	75%	2 µg/L	0.5 µg/L	use best available
Nitrite + nitrate	008	100%	0.072 to 0.3 mg/L	0.4 mg/L	OK
Nitrite + nitrate	009	97%	0.072 to 0.3 mg/L	0.2 mg/L	OK, if best used
Oil and grease	008	21%	0.89 to 0.91 mg/L	~0.1	use best available
Oil and grease	009	32%	0.89 to 0.91 mg/L	~0.1	use best available
Perchlorate	008	35%	0.8 to 1.5 µg/L	0.25 µg/L	use best available
Perchlorate	009	0%	0.8 to 1.5 µg/L	0.25 µg/L	use best available
pH	008	100%	n/a	n/a	OK
pH	009	100%	n/a	n/a	OK
Radium 226+228	008	67%	1.1 to 1.3 pCi/L	0.08 pCi/L	use best available
Radium 226+228	009	75%	1.1 to 1.3 pCi/L	0.08 pCi/L	use best available

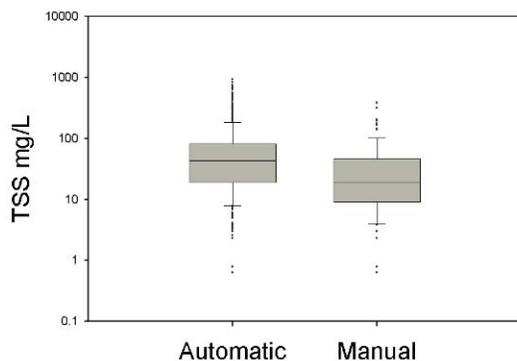
Sulfate	008	100%	0.2 to 2.2 mg/L	3.5 mg/L	OK
Sulfate	009	100%	0.2 to 2.2 mg/L	5 mg/L	OK
TCDD	008	26%	< 1 X 10 ⁻⁹ µg/L	1 X 10 ⁻¹⁰ µg/L	use best available
TCDD	009	65%	< 1 X 10 ⁻⁹ µg/L	1 X 10 ⁻¹⁰ µg/L	use best available
TDS	008	100%	10 mg/L	120 mg/L	OK
TDS	009	100%	10 mg/L	90 mg/L	OK
TSS	008	91%	10 mg/L	10 mg/L	OK
TSS	009	36%	10 mg/L	<1 mg/L	use best available
Zinc, total	008	80%	2.5 to 6 mg/L	10	OK
Zinc, total	009	100%	2.5 to 6 mg/L	5	OK, if best used

Different sampling methods can affect the reported stormwater concentrations. The use of manual or automatic sampling is a factor that is sometimes mentioned as having a possible effect on the quality of the collected samples. Manual sampling is usually used when the number of samples is small and when there are not available resources for the purchase, installation, operation, and maintenance of automatic samplers. Manual sampling may also be required when the constituents being sampled require specific handling (such as for bacteria and for oil and grease) (ASCE/EPA 2002). Automatic samplers are recommended for larger sampling programs, when better representations of the flows are needed, and especially when site access is difficult or unsafe. Automatic samples also improve repeatability by reducing additional variability induced by the personnel from sample to sample (Bailey 1993). Most importantly, automatic samplers can be much more reliable compared to manual sampling, especially when the goal of a monitoring project is to obtain data for as many of the events that occur as possible, and sampling must start near the beginning of the rainfall (Burton and Pitt 2002). Maestre and Pitt (2005) compared data collected from automatic samples and from manual sampling at many NSQD sampling locations in residential, commercial, and industrial locations along the central east coast. One-way ANOVA and Dunnett's test analyses were used to identify any statistical differences between the two groups. From 10 to 200 events were represented in each constituent subset for each sampling method. There were usually more data collected using automatic samplers. As indicated in the following discussion, there were no consistent biases in the data

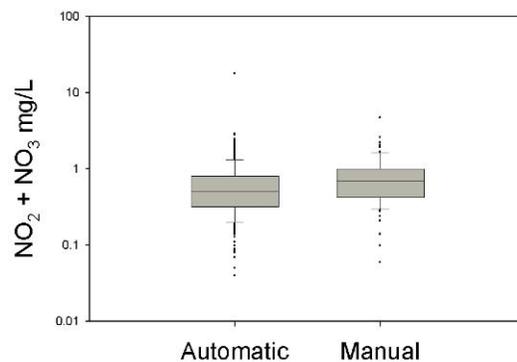
collected by either method. However, automatic sampling has long been regarded as better representing the complete event runoff period. Figure C-2 illustrates some of the comparison plots showing differences in TSS, nitrate+nitrite, Cu and Zn in residential areas. In most cases, the observed differences were relatively small compared to the variability in the data within each subset.

BOD₅ and dissolved phosphorus measurements were not affected by differences in sampling methods used in residential, commercial or industrial areas. In residential and commercial land uses, TSS and COD median concentrations obtained using automatic samplers were almost twice the values obtained when using manual sampling methods. Median total phosphorus concentrations were about 50% higher using automatic samplers, while no effects were noted for other nutrients. TSS, total copper and total zinc had lower concentrations using manual sampling compared with automatic sampling, while the opposite pattern was observed for nitrate-nitrite; manual sampling shows higher median concentrations than samples collected with automatic samples. In industrial land uses, the pattern was found to be opposite. Ammonia, nitrate-nitrite, TKN and total zinc indicated higher median concentrations when using manual sampling methods compared to using automatic samplers. Median concentrations for these constituents were almost twice as high when using manual sampling, except for ammonia that was almost six times higher when manual sampling was used compared to automatic sampling methods. Again, these differences were not consistent and automatic flow-weighted sampling is considered the most representative method for total storm conditions.

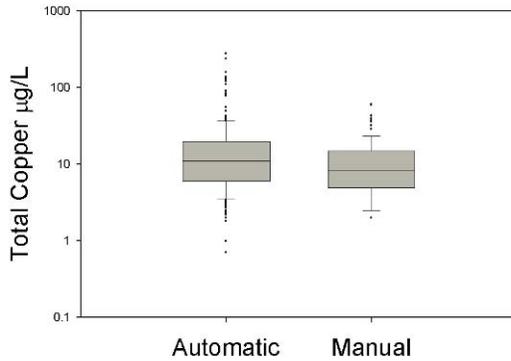
$\alpha < 0.001$



$\alpha = 0.10$



$\alpha < 0.001$



$\alpha < 0.001$

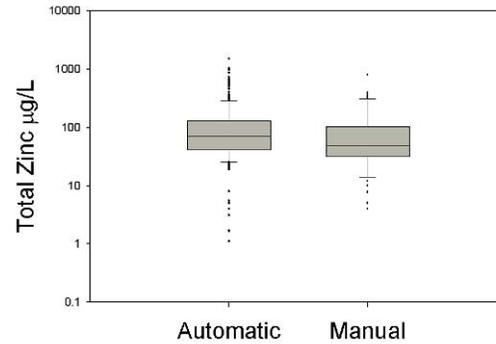


Figure C-2. Comparison of reported concentrations in residential land use for automatic vs. manual sampling methods (mid Atlantic region).

Different sample compositing methods can also affect the stormwater concentrations.

Maestre and Pitt (2005) also compared time and flow-weighted composite automatic sampling options in residential, commercial, and industrial land uses at mid-Atlantic NSQD sites, along with industrial land uses in the southeast. With time-compositing, individual subsamples are combined with even time increments during the runoff event. As an example, automatic samplers can be programmed to collect a subsample every 15 minutes for collection into a large sample bottle. An automatic sampler can also collect discrete subsamples at even time increments, keeping each sample in a separate smaller sample bottle. After the sampled event, these samples can be manually combined as a composite, and if flow data are available, different volumes can be combined from each subsample bottle reflecting the portion of the total flow that occurred during the sampling period to obtain a flow-weighted sample. With automatic flow-weighted sampling, a sampler can be programmed to deposit a subsample into a large sample bottle for each set increment of flow.

The Wisconsin Department of Natural Resources conducted a thorough evaluation of alternative sampling modes for stormwater sampling to determine the average pollutant concentrations for individual events (Roa-Espinosa and Bannerman 1995). Four sampling modes were compared at outfalls at five industrial sites: flow-weighted composite sampling, time-discrete sampling, time-composite sampling, and first flush sampling during the first 30 minutes of runoff. Based on many attributes, they concluded that time-composite sampling at outfalls can be useful due to simplicity, low cost, and good comparisons to flow-weighted composite sampling (assumed to be the most accurate), but only if a very large number of time-composite subsamples are collected. The accuracy and reproducibility of the composite samples were all good, while these attributes

for the first flush samples were poor. Burton and Pitt (2002) stress that it is important to ensure that acceptable time-weighted composite sampling include many subsamples. Any sampling scheme is very inaccurate if too few (and too infrequent) subsamples are collected. Samples need to be collected to represent the extreme conditions during the event, and the total storm duration. Experimental design methods can be used to determine the minimum number of subsamples needed considering likely variations. It is more common to now include the use of “continuous” water quality sondes at sampling locations, with *in-situ* observations of several indicator parameters (pH, ORP, conductivity, turbidity, and temperature) obtained every few minutes.

One-way ANOVA tests were used to evaluate the presence of significant differences between these two composite sampling schemes using these NSQD data. One-way ANOVA with Dunnet’s comparison test was used to evaluate if concentrations associated with time-compositing were larger or lower than concentrations associated with flow-compositing. Figure C-3 is a plot showing the differences between these two sample compositing schemes for TSS, illustrating the large variation in reported concentrations in each subgroup compared to the biases between the compositing methods. The analyses found that no significant differences were observed for BOD₅ concentrations using either of the compositing schemes for any of the four categories based on the number of data observations available (10 to more than 100 observations per category for each constituent). A similar result was observed for COD, except for commercial land uses where not enough samples were collected to detect a significant difference. TSS and total lead median concentrations were two to five times higher in concentration when time-compositing was used instead of flow-compositing. Nutrients collected in residential, commercial and industrial areas showed no significant differences using either compositing method. The only exceptions were for ammonia in residential and commercial land use areas and for total phosphorus in residential areas where time-composite samples had higher median concentrations. Median metal concentrations were higher when time-compositing was used in residential and commercial land use areas. No differences were observed in industrial land use areas, except for lead.

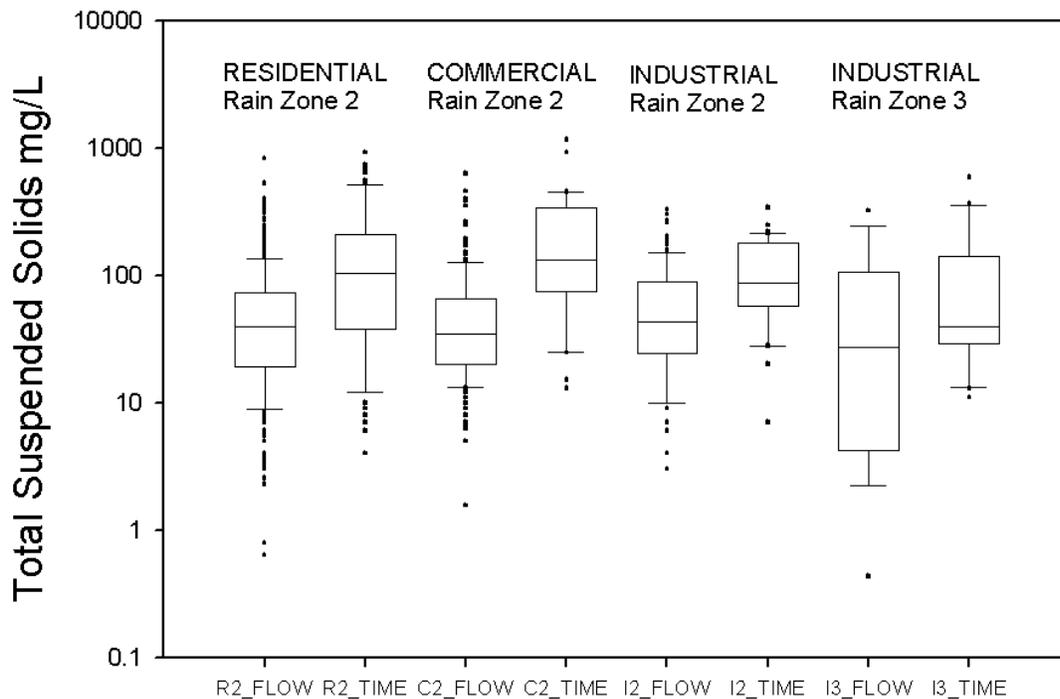


Figure C-3. Comparisons between time- and flow-composite options for TSS (mid-Atlantic and some Southeastern sites)

Short sampling periods during runoff events can affect the reported concentrations. Another potential factor that may affect stormwater quality is the sampling period during the runoff event. Automatic samplers can initiate sampling very close to the beginning of flow, while manual sampling usually requires travel time and other delays before sampling can be started. It is also possible for automatic samplers to represent the complete storm, especially if the storm is of long duration, as long as proper sampler setup programming is performed (Burton and Pitt 2002). The general NPDES stormwater sampling protocols only required collecting composite samples over the first three hours of the event instead of during the whole event. Truncating the sampling before the runoff event ended may have adversely affected the measured stormwater quality, but to a lesser extent than if sampling was only conducted during the first 30 minute period of runoff alone.

The required number of samples to obtain useful information is usually greater than available. Maestre and Pitt (2005) illustrated this point by using data from the Flagstaff Street monitoring station, in Prince George MD (a site having 28 sampled events during 1998 and 1999). A statistical test was made choosing many sets of 6 random events (three for each year) from this set, creating 5,600 different possibilities. Figure C-4 shows the histogram of these

possible results, if only three events per year were monitored for two years. The actual monitored median TSS of the 28 events was 170 mg/L, with a 95% confidence interval between 119 and 232 mg/L. Only 60% of the 5,600 possibilities were inside this confidence interval. About 40% of the possibilities for the observed EMC would therefore be outside the 95% confidence interval for the true median concentration for this low level of sampling. As the number of samples increase, there will be a reduction in the bias of the EMC estimates. In Southern California, Leecaster (2002) determined that ten years of collecting three samples per year was required in order to reduce the error to 10%.

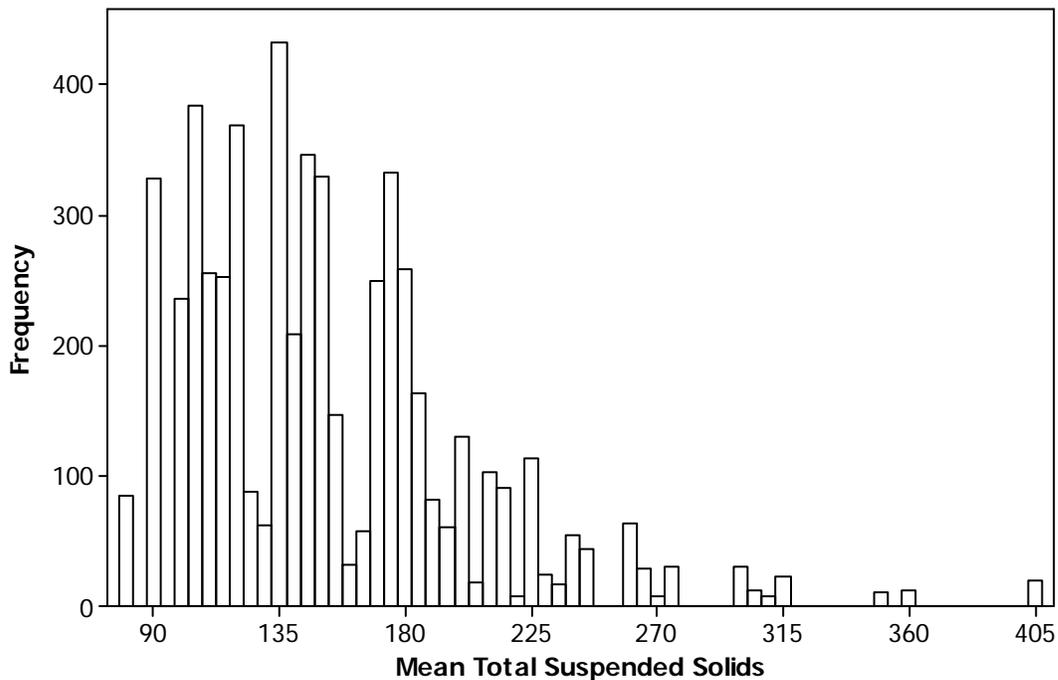


Figure C-4. Histogram of possible TSS concentrations in Flagstaff Street based on collecting three samples per year for two years (the measured median TSS concentration was 170 mg/L)

Selecting a small subset of the annual events can therefore bias the monitoring results. In most stormwater research projects, where stormwater characterization is needed, the goal is usually to sample and analyze as many events as possible covering the complete range of event sizes. As a minimum, about 30 samples are usually desired in order to adequately determine the stormwater characteristics with an error level of about 25% (assuming 95% confidence and 80% power) (Burton and Pitt 2002). With only three events per year required per land use for the NPDES stormwater permits, the accuracy of the calculated EMC is questionable until many years have passed.

As noted previously, probability plots of the available data from Santa Susana outfalls 008 and 009 were prepared to estimate the variability of the values and to determine the likely numbers of samples needed before confident statistical findings can be determined. There were about 20 events at outfall 008 and about 30 events at outfall 009 that were monitored during the 40 months of monitoring. Not all events had samples analyzed for all constituents and not all samples had detected observations. Table C-2 lists the number of samples analyzed for each constituent, along with the predicted number of samples that may be desired. Because of the relatively few runoff events that occur at the monitored outfalls (<1 per month), there are few data available for many of the constituents, even with more than three years of monitoring every runoff opportunity. Coupled with the relatively high variability of the stormwater (not unusual), these results need to be carefully interpreted. The number of data needed for more confident results (considering appropriate power) can be much larger than the current data set for almost all of the constituents. Variability values were calculated based on the slopes of the probability plots. This procedure is described in Burton and Pitt (2002), along with methods used to calculate the numbers of needed samples based on different data quality objectives. The data quality objectives considered in these calculations were 25% errors, confidence of 95%, and power of 80%.

Table C-2. Stormwater Characteristics as Observed at Outfalls 008 and 009

Constituent	Outfall	samples obtained	approximate # of samples needed (25% error, 95% confidence, 80% power)
Antimony, total	008	19	25
Antimony, total	009	31	50
Antimony, filtered	008	3	5
Antimony, filtered	009	9	25
Boron	008	3	20
Boron	009	4	25
Cadmium, total	008	19	~100
Cadmium, total	009	31	~100
Chloride	008	19	40
Chloride	009	31	50

Copper, total	008	19	45
Copper, total	009	31	50
Copper, filtered	008	3	10
Copper, filtered	009	9	10
Gross alpha radioactivity	008	5	40
Gross alpha radioactivity	009	7	~100
Gross beta radioactivity	008	5	50
Gross beta radioactivity	009	7	50
Lead, total	008	19	~100
Lead, total	009	31	~100
Lead, filtered	008	3	n/a
Lead, filtered	009	9	~100
Mercury, total	008	19	40
Mercury, total	009	31	40
Nickel, total	008	3	15
Nickel, total	009	4	~100
Nitrite + nitrate	008	19	50
Nitrite + nitrate	009	31	60
Oil and grease	008	19	60
Oil and grease	009	31	80
Perchlorate	008	20	45
Perchlorate	009	6	45
pH	008	15	5
pH	009	26	5
Radium 226+228	008	3	75

Radium 226+228	009	4	75
Sulfate	008	19	40
Sulfate	009	31	70
TCDD	008	19	~200
TCDD	009	31	~200
TDS	008	19	25
TDS	009	31	25
TSS	008	11	~100
TSS	009	22	~100
Zinc, total	008	5	25
Zinc, total	009	4	65

Automatic samplers may not represent the complete range of particle sizes. Automatic samplers are not capable of sampling bed load material, and are less effective in sampling larger particles. They can be effective in representing particles up to about 250 μm if the sampler intake is suitably located to collect subsamples from a well-mixed sample, such as from a cascading stream (Clark, *et al.* 2008). Manual sampling, also if able to collect a sample from a well-mixed flow (or depth-integrated), can represent the complete particle size distribution. Bed load samples and special floatable capture nets may be needed to supplement automatic samplers if information for the complete range of solids is needed.

Recommended Sampling Approaches to Reduce Possible Problems

The following recommendations were made by Maestre and Pitt (2005) based on evaluating the NSQD data for future stormwater permit monitoring activities:

- The use of automatic samplers, coupled with bedload samplers, is preferred over manual sampling procedures. In addition, flow monitoring and on-site rainfall monitoring needs to be included as part of all stormwater characterization monitoring. The additional information associated with flow and rainfall data will greatly enhance the usefulness of the much more expensive water quality monitoring. Flow monitoring must also be correctly conducted, with adequate calibration and correct base-flow subtraction methods applied. A related issue frequently mentioned by the monitoring agencies is the lack of on-site rainfall information for many of the sites. Using regional rainfall data from locations distant from the monitoring location is likely to be a major source of error when rainfall factors are being investigated.

- Many of the stormwater permits also only required monitoring during the first three hours of the rain event. This may have influenced the event mean concentrations if the rain event continued much beyond this time. Flow-weighted composite monitoring should continue for the complete rain duration.
- Monitoring only three events per year from each monitoring location requires many years before statistically adequate numbers of observations are obtained. In addition, it is more difficult to ensure that a small fraction of the total number of annual events is representative. Data quality objectives must be established and necessary detection limits to control the level of non-detectable observations to acceptable levels (usually to <10%, at least) are needed. Suitable numbers of monitored events over the complete range of rainfall conditions are also needed.

Sampling Approaches to Enable Monitoring of a Wide Range of Rain Events and to Examine Variations in Water Quality during Runoff Events

Automatic samplers can operate in two basic sampling modes, based on either time or flow increments. The sample bases can generally hold up to 24 bottles, each 1 L in volume. A single sample bottle of up to about 20 L is generally available for compositing the sample into one container. These basic bottle choices and the cycle time requirements of automatic samplers restrict the range of rain conditions that can be represented in a single sampler program for flow-weighted sampling. It is important to include samples from the smallest rains that can produce runoff (as small as about 0.01 to 0.2 inch for paved areas, to about 0.5 inches for mostly undeveloped areas) in a stormwater sampling program because they are frequent. Moderate sized rains (from about 0.2 to 3 inches) are very important because they represent the majority of flow (and pollutant mass) discharges for most areas. The largest rains (greater than about 3 inches) are also important because when they occur (infrequently), they can produce large amounts of runoff. It is very difficult to collect samples representing a wide range of rain depths in an automatic sampler using a conventional flow-weighted sampling. Conflicts occur between needing to have enough sub-samples during the smallest event desired (including obtaining enough sample volume for the chemical analyses) and the resulting sampling frequency during peak flows for the largest sampling event desired.

This problem was solved during numerous stormwater monitoring projects (including Pitt and Shawley 1982 during the Castro Valley, CA NURP project, and Pitt 1985 during the Bellevue, WA NURP project, as shown in Figure C-5) by substituting a large container for the standard sample base and installing the sampler in a small shelter. The large container can be a large steel drum (Teflon™ lined), a stainless steel drum, or a large Nalgene™ container, depending on the sample bottle requirements. In order to minimize handling the large container during most of the events, a 10 L smaller container can be suspended inside to collect all of the sub-samples for the

majority of the events. The jar would overflow into the large container for the largest events. The small shelter should be well vented to minimize extreme temperatures, as it is difficult to ice the large container. Obviously, the sampling stations need to be visited soon after a potential runoff event to verify sample collection, to collect and preserve the collected sample, and to clean the sampler and prepare it for the next event.

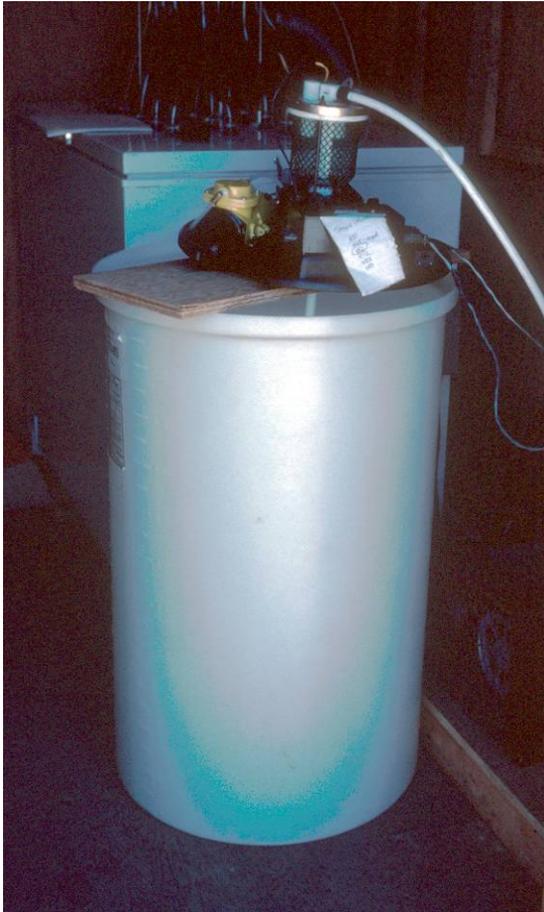


Figure C-5. Automatic sampler with large base for monitoring wide range of flows, with large chest freezer USGS discrete sampler in background, at Bellevue, WA (Pitt 1985)

Alternatives to using a large sample base in order to accommodate a wide range of runoff events include:

- Use time-compositing instead of flow-weighted sampling and then manually composite the sample using the available flow data (described below),
- Use two samplers located at the same location, one optimized for small events, the other optimized for larger events (Figure C-6), or

- Visit the sampling station during the storm and re-program the sampler, switch out the bottles, or manual sample.



Figure C-6. Double monitor setup for sampling over a wide range of flow conditions at Tuscaloosa, AL (Pitt, current projects)

The most common option is the last one which is expensive, uncertain, and somewhat dangerous. Few monitoring stations have ever used multiple samplers, but that may be the best all-around solution, but at an increased cost. One sampler would be programmed for the smaller events, while the other sampler would have a complementary program emphasizing larger events.

In recent years, new sampler options have become available. The ISCO STORM programming option for the model 6712 samplers is briefly described below that enables combinations of these options to be used, along with capturing a separate sample in the first 30 minutes of the event. One flexible arrangement is illustrated in Figure C-7 with twelve 1-quart glass bottles in the sampler base (up to 24 plastic containers can also be used). The initial three bottles are filled in 10 minutes each and individual samples are collected into the nine remaining bottles at 20-minute intervals (the multiplexer option can also be used to fill multiple subsamples in each

bottle). After the storm, sub-samples from each of these nine bottles can be poured into one container, with the volumes taken from each bottle representing the flow increment during that sampling period. This would result in a flow-composite sample in addition to an initial 30 minute sample. In addition, the multiple bottles can be individually analyzed for key indicator constituents if concentration variations during the event are desired.

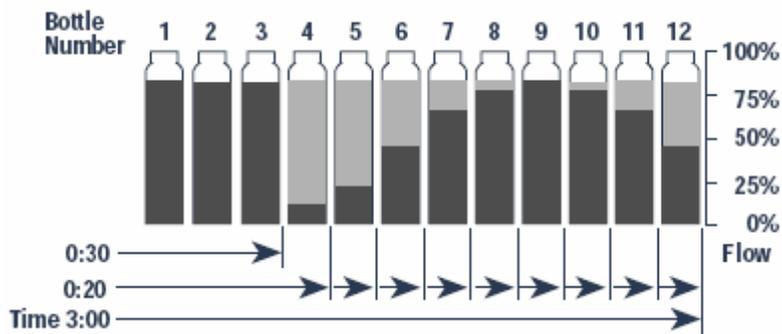


Figure C-7. Sampler schematic for first flush and flow-weighted samples (ISCO)

This last option requires a lot of manual processing of the samples and is only suitable when a limited number of sampling stations are being monitored. A more effective approach is to use automated methods as previously described (enlarged containers or double samplers). If concentration variations are desired during the event, it is recommended that continuous water quality sondes be used. These sondes can also be used to trigger sampling if unusual constituent concentrations are observed (mostly used for continuous point source discharges when upset conditions are being investigated, and when relationships between the indicator parameters monitored by the sonde and the constituents of concern are significant).

YSI 6000 series sondes are capable of continuously monitoring pH, ORP, temperature, conductivity, temperature, turbidity, DO, and water depth. We have used these instruments in continuous 30 to 45 day deployments in monitoring urban receiving waters and stormwaters with minimal drift in the calibrations. These sondes can also be fitted with specific ion electrodes, but these have not worked well under the harsh conditions encountered. We program the sondes to obtain readings every 5 minutes when a runoff event is expected. When the rain ends, but if flows may still occur for an extended period, we re-program them to obtain data every 15 minutes. This enables the battery life and the data storage capacity to extend for several weeks. We re-calibrate the sondes before the next event is expected, obtaining an indication of drift since the last calibration and conducting any needed maintenance (such as replacing the batteries). Figure C-8 is an example of the basic sonde data analysis screen. Detailed spreadsheet data logs are downloaded from the sondes to enhance extended analyses. This information allows us to examine basic water quality variations during the sampling period.

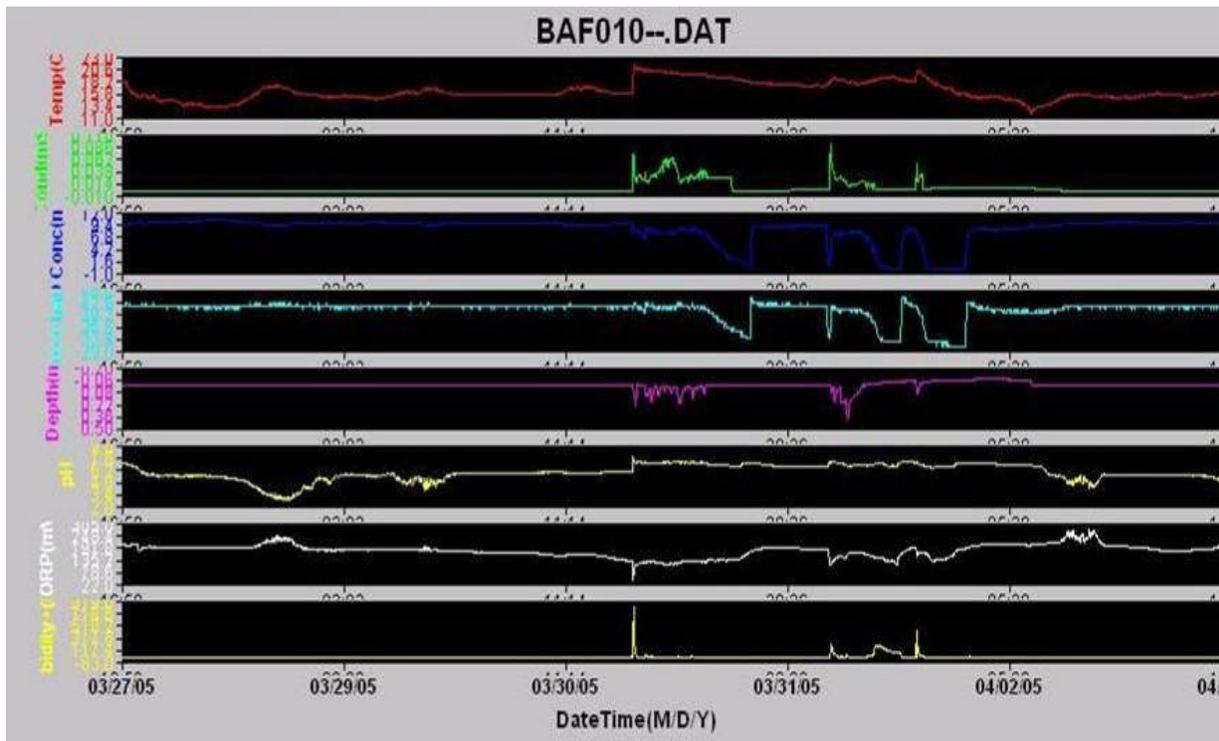


Figure C-8. Screen shot of sonde data analysis screen showing ten days of high-resolution water quality measurements (Pitt and Khambhammettu 2006).