TECHNICAL PAPER #0607

STORMWATER MANAGEMENT PROGRAM LONG-TERM EFFECTIVENESS MONITORING:

Can I Reduce Sampling Frequency and Still Detect Trends to Determine BMP Effectiveness?

Prepared by:

David Pohl, Ph.D.
Susan Watts
Weston Solutions, Inc. (WESTON®)

Bill DePoto
Los Angeles County Department of Public Works

For Presentation at:

2006 StormCon

24 - 27 July 2006
Denver, CO, USA

STORMWATER MANAGEMENT PROGRAM LONG-TERM EFFECTIVENESS MONITORING – CAN I REDUCE SAMPLING FREQUENCY AND STILL DETECT TRENDS TO DETERMINE BMP EFFECTIVENESS?

Bill DePoto, P.E. (Los Angeles County Department of Public Works), David Pohl, P.E., Ph.D., and Susie Watts (Weston Solutions, Inc.)

INTRODUCTION

Stormwater program managers are challenged with meeting regulatory monitoring requirements to identify long-term trends and the effectiveness of the implementation of Best Management Practices (BMPs) and meeting multiple program requirements with finite resources. This paper provides stormwater managers with methods and several real-world scenarios using 10 years of data from the Los Angeles County Stormwater Monitoring Program to design long-term effectiveness monitoring programs. Potential tradeoffs in frequency reduction of sampling are assessed with increased focused monitoring of smaller drainage areas where BMPs have been implemented to optimize available resources.

The paper first presents the results of statistical analysis of up to 10 years of stormwater monitoring results collected by the Los Angeles County Department of Public Works for its Core Monitoring Program as specified in the Municipal Stormwater Permit. The data set represents runoff within a receiving channel from a highly urbanized watershed. The statistical analysis includes the evaluation of the inherent variability of urban runoff and the number of samples required to determine a significant difference for various confidence levels. The statistically required number of samples is determined for a selected metal (total copper) compared to suggested sample sizes in published guidelines to determine effectiveness of BMPs. The variability of the data set from the in-channel monitoring station at the base of the watersheds is also compared to the water quality variability of a sub-drainage area within the watershed. The findings of these evaluations are summarized to provide stormwater managers with the basis to plan and design BMP effectiveness assessment monitoring programs.

The paper also presents the results of the statistical evaluation using the Los Angeles County data of modifying sample frequency for determining long-term trends in stormwater quality. Several real-world scenarios are presented showing trends detected by different sample frequencies for different reductions in constituent concentrations through the implementation of BMPs. The results presented provide stormwater managers with methods and statistically based potential tradeoffs in monitoring programs to assess long-term effectiveness of stormwater management programs.

ASSESSMENT OF SAMPLING PROGRAMS TO VERIFY BMP EFFICIENCIES

Stormwater runoff from urbanized watersheds can be highly variable due to multiple land uses and constituent sources. In highly urbanized areas of southern California like Los Angeles, stormwater is conveyed from urban landscapes to storm drains and into channelized receiving waters, reducing the ability of the natural environment to attenuate peak flows and constituent loadings. These urbanized conditions can result in high variability in constituent loadings depending on the storm event size and duration. Best Management Practices (BMPs) are being implemented in these urban settings to reduce peak flows and constituent loadings in order to restore and maintain designated beneficial uses of receiving waters. As BMPs are installed, their effectiveness is being investigated to confirm design removal efficiencies and to determine if further BMPs are needed to reduce loadings.

BMP removal efficiencies are measured by comparing constituent loading of the stormwater inflow into the structural BMP, or at an up-stream location from a drainage area where nonstructural BMPs have been implemented, to the reduction in loading from the effluent or downstream location. To determine the number of samples needed to verify that the design load reduction has been achieved, a statistical power analysis was conducted using 10 years of stormwater data. The data set used for this statistical analysis was from the mass emissions sampling station on the Los Angeles River and on Ballona Creek. The samples from the mass emission stations are collected using flow-weighted composite sampling and therefore are representative of event mean concentrations (EMCs). Both of these receiving waters are located in urbanized watersheds in Los Angeles County. More than half of the Los Angeles River Watershed is developed and approximately 32% is impervious area based on land-use types. Approximately 40% of the Ballona Creek Watershed is estimated to be impervious. Land-use distribution for the Los Angeles River and Ballona Creek Watersheds is shown on Figures 1 and 2.

The power analysis was conducted for three design load reduction efficiencies of 20%, 30%, and 40%. The results of the power analysis using the Los Angeles River and Ballona Creek data are presented on Figures 3 and 4, respectively. These data sets consist of approximately 20 samples collected over a 6- to 10-year period. The power is defined by the probability of a type II error, or the probability of rejecting a null hypothesis when it is false. For example, a 0.8 power indicates that the design reduction would be detected 80% of the time. Figures 3 and 4 present the number of samples and the corresponding power for the various percent reductions assumed for the BMP or series of BMPs. The results of this analysis indicate the following:

- Greater load reductions are required to achieve higher power. For both the Los Angeles River and Ballona Creek Watersheds data, higher powers were achieved for the 40% reductions compared to the 20% and 30% reductions.

- Greater variability in the data reduces the power level. Ballona Creek is characterized by greater variability than the Los Angeles River. The power level for Ballona Creek, even at greater reductions, was half that of the Los Angeles River.

- Approximately 15 to 20 samples are needed to confirm the design reductions and achieve the optimum power level.

- For the two sets of data from the mass emission station at the base of the watershed, increased numbers of samples do not significantly change the power after approximately 25 samples. The percent reduction is equally applied to the variability due to error. As a result, greater percent reductions will also reduce overall variability, including error. This effectively reduces the chance for a type II error and increases power.

- Due to the high variability of stormwater data, power levels of 0.40 may be the highest achievable level, as indicated for the Ballona Creek watershed management area (WMA) data.

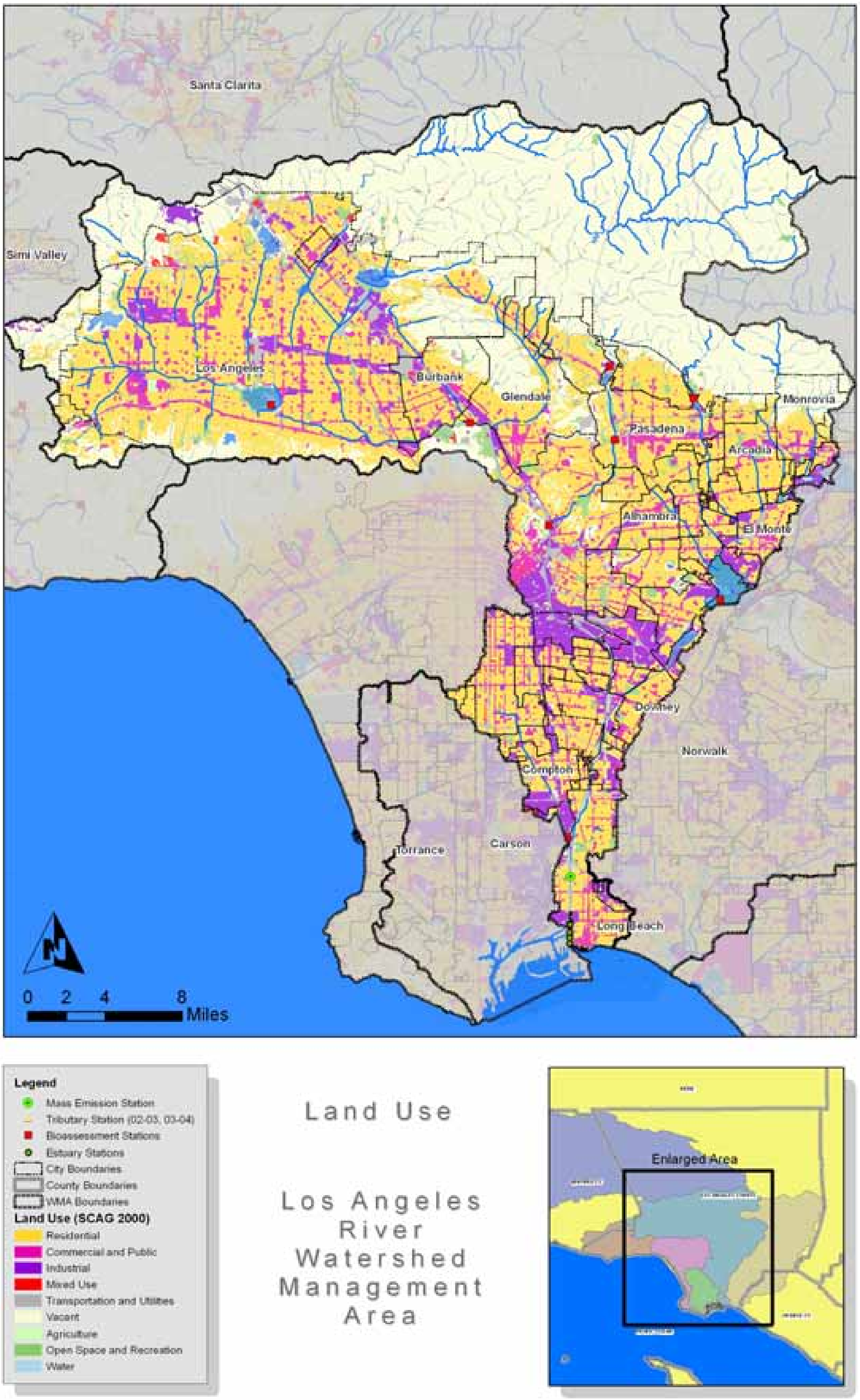

Figure 1. Land Use Distribution in the Los Angeles River Watershed.

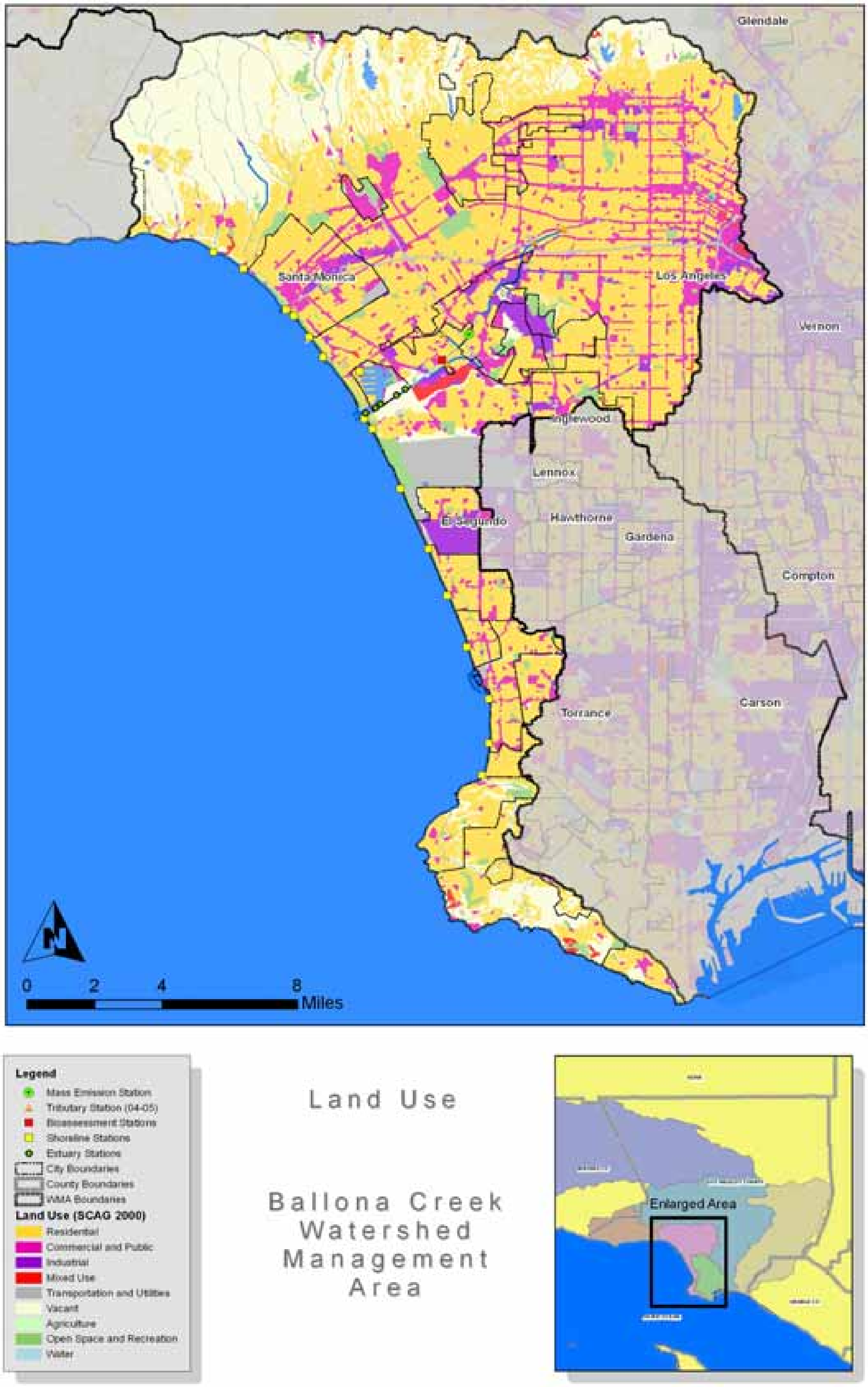

Figure 2. Land Use Distribution in the Ballona Creek Watershed.

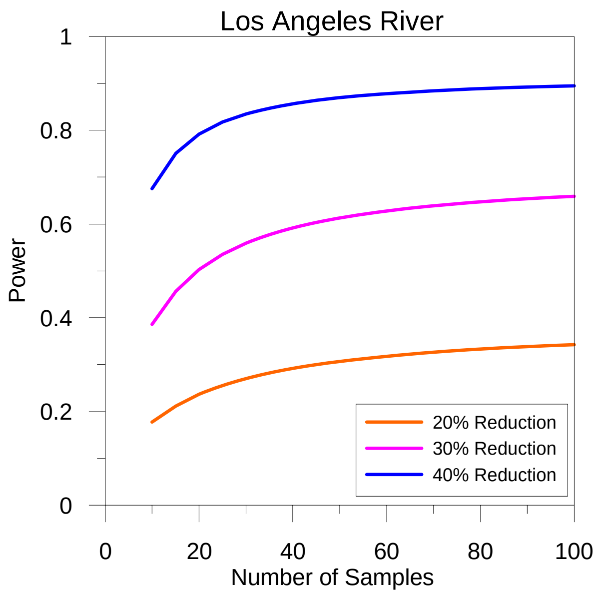

Figure 3. Power Analysis – Verification of BMP Removal Efficiency – Los Angeles River Watershed.

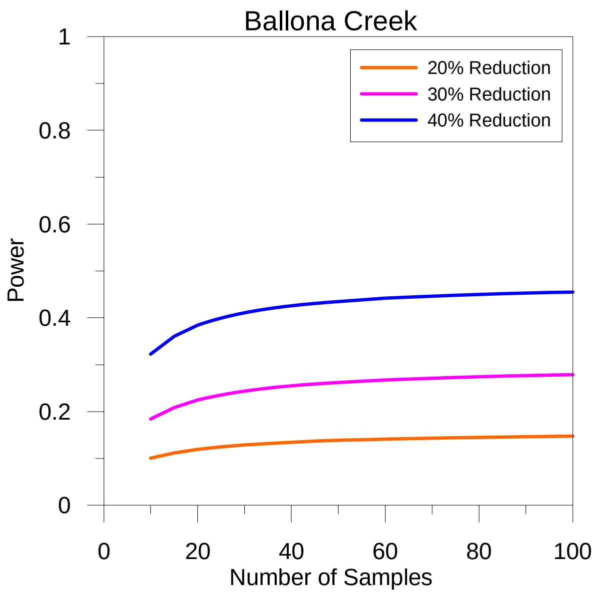

Figure 4. Power Analysis – Verification of BMP Removal Efficiency – Ballona Creek Watershed.

As stated above, the power analysis curves shown on Figures 3 and 4 for the data from the mass emission stations are characterized by a flattening of the curves (no significant change in power) after 25 samples.

The above analysis was conducted using stormwater data that represents urban stormwater runoff from a large urbanized watershed. To assess the characteristics of stormwater from a smaller drainage area that may represent typical inflows to an individual BMP or smaller sub-watershed where nonstructural BMPs are being implemented, the results from flow-weighted composite sampling from a tributary of the Ballona Creek Watershed was analyzed. This data set consists of five samples. The results of the power analysis for this data set are presented in Figure 5.

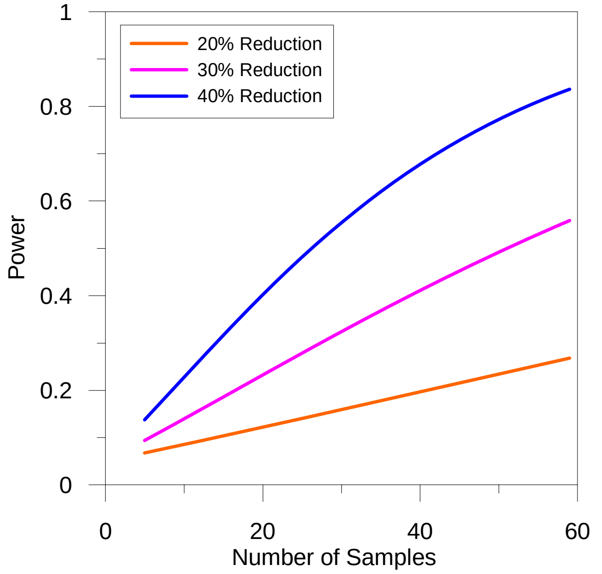

Figure 5. Power Analysis – Verification of BMP Removal Efficiency – Sub-Watershed – Tributary to Ballona Creek.

As shown on Figure 5, the smaller drainage area results indicate a greater sensitivity to sample number than the watershed results presented previously. The percent reduction still greatly influences the power attained, but the power curves for the smaller drainage area do not flatten out after 25 to 30 samples, like the larger watersheds. The factors that influence the shape of the curve include both variability and size of the original data set. The smaller drainage area exhibited overall greater variability, but this is in part due to the smaller sample size. The results are consistent with the previous curves that indicate approximately 20 samples at reductions greater than 30% are needed to achieve a power of 0.5 or greater. For lower reductions, event at large sample sizes, the power is still below 0.25.

The results of the power analysis indicate that approximately 20 samples (storm events) may be needed to achieve an appropriate power to verify the effectiveness of specific BMPs or a series of BMPs applied within a watershed or drainage area. This number compares to the 15 to 20 sampling events recommended in the Protocol for Stormwater Best Management Practice Demonstrations (The Technology Acceptance Reciprocity Partnership [TARP], July 2003). This recommended sample number is, however, greater than what is typically scoped for BMP effectiveness programs due to available resources. A tradeoff is therefore needed to balance an increased effort in BMP effectiveness monitoring with other monitoring activities, such as the frequency of long-term water quality trend analysis sampling. The following statistical analysis will evaluate potential approaches to identifying tradeoffs using finite available resources.

EVALUATION OF POTENTIAL REDUCTIONS IN THE FREQUENCY OF LONG-TERM TREND SAMPLING

Should increased sampling be needed to verify the effectiveness of BMPs implemented in various drainage areas within a watershed as indicated in the above analysis for a highly urbanized watershed, stormwater managers will need to assess potential tradeoff in order to meet the overall requirements of their stormwater programs within the finite available resources. A potential tradeoff may include the reduction in the frequency of long-term water quality sampling programs that monitor trends for the entire watershed. The recommendation to change the quantity or frequency of sampling therefore necessitates a thorough review of the impacts on following established trends in the data and the ability to detect changes or trends in the data. To determine the impacts of reducing the quantity and/or frequency of sampling the composite samples collected at the base of the watershed, an analysis was performed to simulate future data and determine the frequency at which sampling would need to be performed to ascertain with 95% confidence that concentrations were below the pertinent water quality objective (WQO).

The analysis uses $\log_{10}$ transformed data to determine the equation of the regression line drawn through the existing data. Log transformation is commonly used to normalize data when concentrations are of different orders of magnitude. The regression equation is used to compute the predicted mean value in future years and the standard deviation from the regression analysis. Data are generated for each future year using a Monte Carlo statistical approach. Using a mean equal to the predicted mean, the data are adjusted for a specific reduction, and finally randomly distributed within the bounds of the standard deviation. Using these simulated data for the various percent reductions with the existing data, the regression is rerun and the point in time when the upper 95% confidence bound of the regression crosses below the WQO is determined. Because this is just one random simulation that may be anomalous, this process is repeated with 100 sets of randomly simulated data based on the original equation. The entire set of 100 regressions is then evaluated to determine when the upper confidence bound is below the WQO 95 out of 100 times. This is the number of years of sampling that must occur to be confident that the concentration meets the objective. This whole process was performed for sampling a) three storm events every year into the future, b) three sampling events in alternate years, c) two storm events every year, and d) two storm events in alternate years.

The results of this statistical analysis are presented in Figures 6, 7, and 8 for estimated reductions due to BMPs of 20%, 30%, and 40%, respectively, using 10 years of stormwater data from the Los Angeles River. Presented on these figures is a comparison of the year that would be anticipated to reach the WQO for total copper for different sampling frequencies. Each of the following figures for the three different levels of loading reductions contain four separate timelines when the regression line (blue line) and the 95% confidence level line (green line) intersect the WQO concentration for three events per year, three events per year using alternate years, two events per year, and two events per year for alternate years. The vertical lines represent the year when the WQO is met. The regression line and 95% confidence line were generated assuming the estimated reductions of 20%, 30%, and 40%.

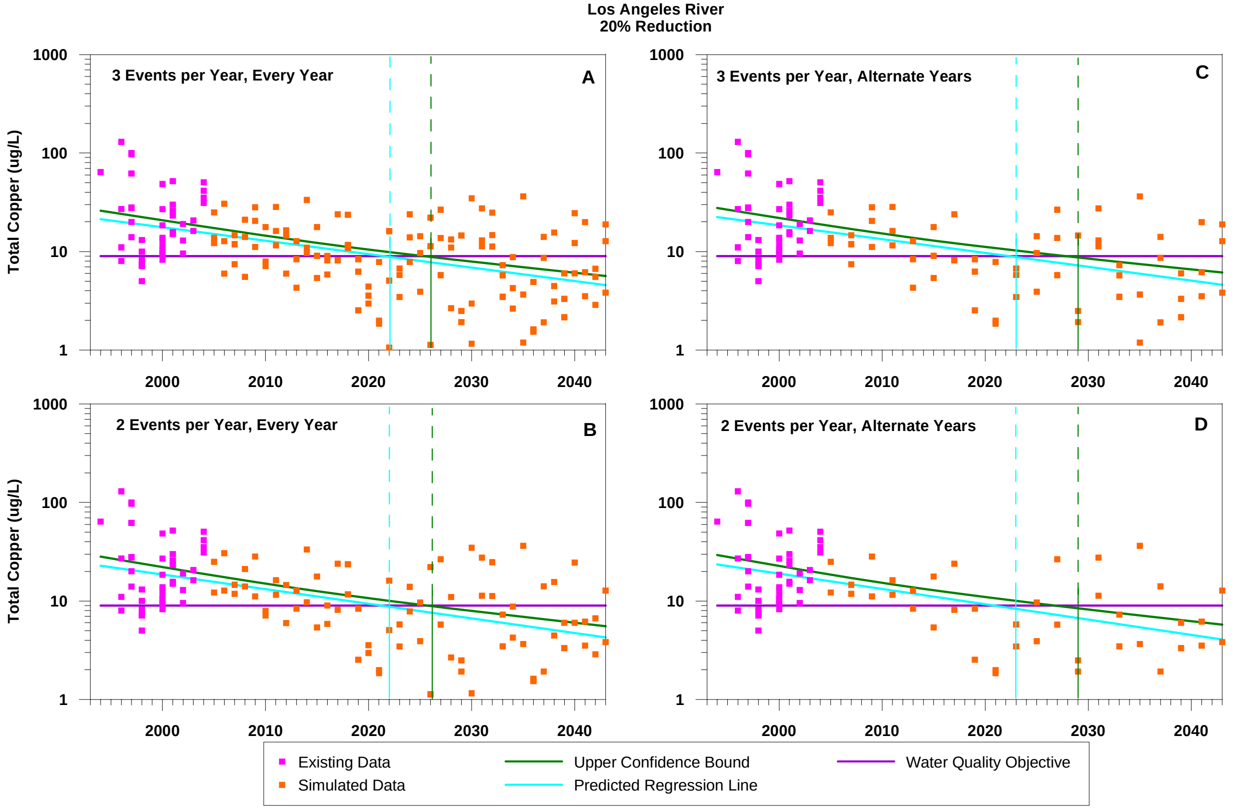

Figure 6. Regression Analysis for Total Copper – Reduction of 20%.

Figure 7. Regression Analysis for Total Copper – Reduction of 30%.

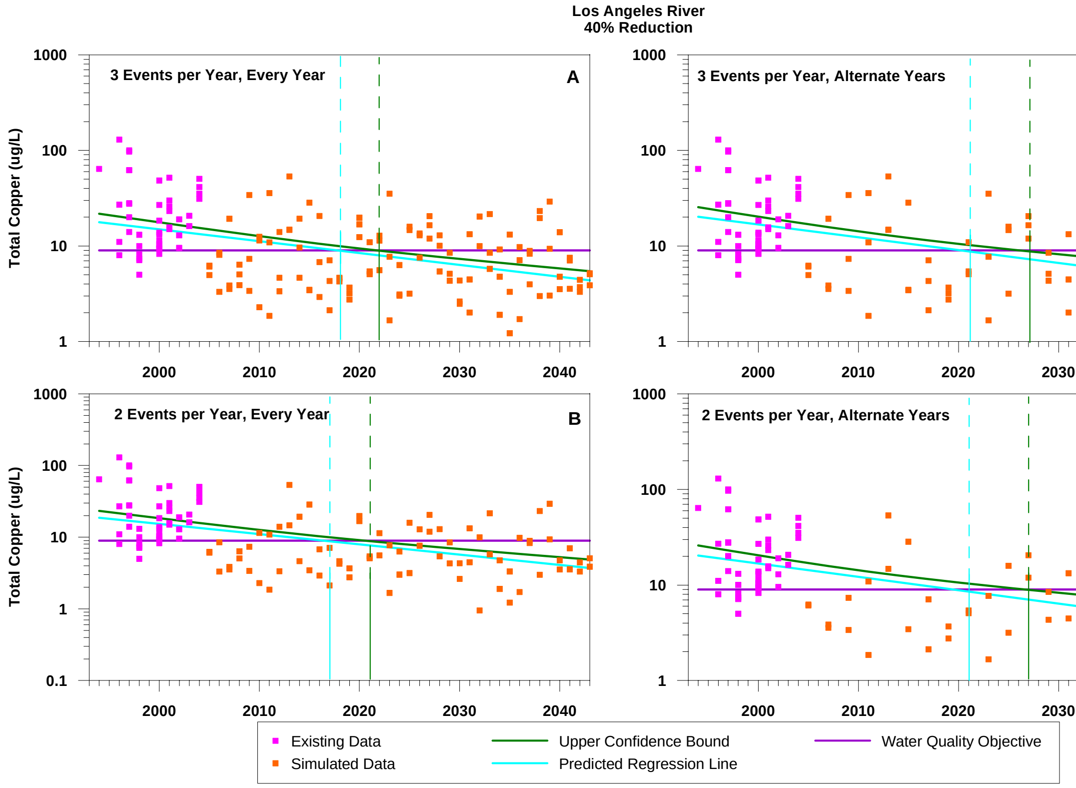

Figure 8. Regression Analysis for Total Copper – Reduction of 40%.

As shown in Figure 6, for an estimated trend indicating a 20% reduction, reducing the frequency of sampling from three events every year to two events every year does not change the predicted time when regression or upper confidence bound intersects with the WQO line. Therefore, a change in sampling from three to two events annually will not impact a stormwater manager's ability to identify trends and completion of objectives through the implementation of BMPs that result in a 20% reduction in loadings. The analysis for the reduction of sample frequency to three and two events on alternating years indicates no difference between the three and two events on this reduced timeframe. The results also show that the trends and time of detection when the objective is achieved can be determined within a typical 5-year permit cycle event at these reduced sampling frequencies compared to the three sampling events annually. Although the time that the objective is met is pushed out several years, it is still within this 5-year permit period. Therefore, reductions in sampling frequency can be implemented, while still meeting requirements to track long-term trends and detect when the applicable WQOs are met. Similar results are also indicated for reduction trends of 30% and 40%. At greater reductions (slope of the regression line is steeper), the difference between the dates when the regression line intersects the WQO line is greater between the three events per year and the two events in alternating years. However, the difference is still within a 5-year timeframe. Therefore, a lower frequency than that currently used for monitoring will still be effective in assessing water quality of receiving waters from the entire monitored watershed and detecting the reduction of key constituents down to the WQO within a similar timeframe.

CONCLUSIONS

The results of the statistical analysis of the 10 years of stormwater data from Los Angeles County confirm a high degree of variability in constituent concentrations in urban runoff. This variability can be expected due to multiple land uses and constituent sources. Stormwater managers challenged with verifying the design effectiveness of implemented BMPs need to consider this inherent variability and the subsequent level of sampling effort required to meet minimum statistical tests. The statistical power analysis presented for two urban rivers indicated that approximately 15 to 20 samples may be needed to verify effectiveness of a BMP or a series of BMPs that are designed to reduce loadings in urban stormwater runoff. This is consistent with recommendations in the Protocol for Stormwater Best Management Practice Demonstrations (TARP, July 2003). The analysis also demonstrated that to achieve a power of greater than 0.5, and therefore confirm that greater than 50% of the results achieve the design reduction, load reductions of 30% and higher are required. Even at greater numbers of samples, the power does not increase above 0.2 to 0.3 for systems with lower reductions. For the data collected from the mass emissions stations located at the base of the watershed, the power curves flattened out at 20 to 25 samples, indicating percent reduction has much greater influence than number of samples for this data set. In comparing these results with the stormwater data of a smaller drainage area within the larger watershed, the data are characterized by greater variability. For the smaller drainage area, the power increases with the increased number of samples. However, at percent reductions less than 40%, a power of 0.5 is not achieved for sample numbers less than 50, and are never attained for the 20% reduction. Comparisons of inflow and outflow loadings from systems that achieve lower reductions will likely not be practical, and therefore effectiveness sampling should be conducted where higher reductions are anticipated.

In addition to verifying the effectiveness of specific BMPs or a series of BMPs in specific drainage areas, stormwater managers are also challenged with tracking trends in the water quality of the receiving waters compared to WQOs. To assess if tradeoffs could be made to reduce frequency of sampling at the base of the watershed to provide resources up into the watershed to verify BMP design effectiveness, a regression analysis of the stormwater results from the Los Angeles River was conducted using a Monte Carlo statistical approach. The results of this analysis indicated that trends and verification of meeting the WQO could still be detected within the 5-year permit cycle using reduced sample frequency. Reductions from three to two times annually or using an alternating-year approach still allow for detection of the trends at meeting the objective within the permit cycle. This was the case for the various percent reductions evaluated (i.e., 20%, 30%, and 40%).

These results are based on assumed reductions and statistical methods that randomly select future concentrations using characteristics of the existing data. The results therefore do not represent actual measured trends, but can be used as a tool for stormwater managers to develop long-term effectiveness assessment monitoring programs and utilize finite resources effectively. This analysis focused on measuring BMP effectiveness through comparisons of inflow and outflow EMCs. Other questions and subsequent monitoring that need to be performed to assess BMP effectiveness may include reduction in toxicity in the receiving waters, and improvement/protection in downstream biotic communities and downstream physical habitat impacts. Finally, the statistical analysis performed should be verified using data from other subject watersheds and using nonparametric trend methods based on the distribution of the stormwater data.