Assessment of efficient sampling designs for urban stormwater monitoring

Molly K. Leecaster[1], Kenneth C. Schiff, and
Liesl L. Tiefenthaler

ABSTRACT

Monitoring programs for urban runoff, which are highly variable and do not fit a point source model, have not been assessed for effectiveness or efficiency in estimating mass emissions. In order to determine appropriate designs for stormwater, total suspended solids (TSS) and flow information from the Santa Ana River was collected nearly every 15 min for every storm of the 1998 water year. All samples were used to calculate the "true load" and then three within-storm sampling designs (flow-interval, time-interval, and simple random) and five among-storm sampling designs (stratified by size, stratified by season, simple random, simple random of medium and large storms, and the first *m* storms of the season) were simulated. Using these designs, we evaluated three estimators for storm mass emissions (mean, volume weighted, and ratio) and three estimators for annual mass emissions (median, ratio, and regular). Designs and estimators were evaluated with respect to accuracy and precision. The optimal strategy was used to determine the appropriate number of storms to sample annually based upon confidence interval width for estimates of annual mass emissions and concentration. The amount of detectable trend in mass emissions and concentration was determined for sample sizes 3 and 7. Single storms were most efficiently characterized by taking 12 samples following a flow-interval schedule and using a volume-weighted estimator of mass emissions. This design and estimator had the best combination of small bias and standard error. Randomly selecting the medium and large storms within a season achieved the smallest bias for concentration and reasonable bias for estimating mass emissions. This design also attained a small standard error. The ratio estimator most accurately estimated concentration and mass emissions from the simple random sample of medium and large storms, and had low bias over all of the designs. This estimator minimized standard error when coupled with the simple random sample of medium and large storms. Sampling seven storms is the most efficient method for attaining small confidence interval width for annual concentration. Sampling three storms per year allows a 20% trend to be detected in mass emissions or concentration over five years. These results are decreased by 10% by sampling seven storms per year.

INTRODUCTION

Urban runoff is a large source of mass emissions to coastal oceans (Schiff and Tiefenthaler, 2001). Runoff contains pollutants that pose a risk to human health (Haile *et al.* 1999) as well as to indigenous plants and animals (Bay and Schiff 1997). This risk is compounded in southern California where most watersheds are highly developed and precipitation is infrequent, which may result in an increase in the number of sources and pollutant accumulation over longer periods of time prior to highly variable seasonal flows.

Routine monitoring of urban runoff discharges is in its early stages of development, and little consistency or comparability has been achieved among monitoring programs (Schiff 1997). This problem is further compounded by the absence of testing programs to evaluate urban runoff sampling strategies for effectiveness and efficiency; therefore, an optimal program has not been identified. Existing monitoring programs for point sources are inappropriate since stormwater flows and concentrations vary by orders of magnitude in a matter of hours (Cross *et al.* 1992).

The objective of this study is to assess various urban stormwater sampling designs. Data from a comprehensively measured system were subsampled to simulate various strategies, and estimators of mass emissions and

---

[1] *Present address: INEEL, Bechtel WBXT, Idaho, LLC, P.O. Box 1625, Idaho Falls, ID 83415-3779*

concentration of total suspended solids (TSS) were compared.

METHODS

Sampling strategies that characterize runoff from the Santa Ana River were evaluated using Monte Carlo simulations taken from a year of continuous stormwater sampling. Three strategies were evaluated for their effectiveness in sampling within storms and five strategies were evaluated for their ability to select storms to sample (among-storm sampling). Three estimates of single-storm mass emissions and concentration and three estimates of annual mass emissions and concentration were considered. Optimal monitoring strategies and estimators were chosen to maximize the accuracy and precision of mass emissions and concentration of TSS. The optimal strategy was then used to determine the appropriate number of storms to sample per year based upon confidence interval width for annual mass emissions and concentration. The amount of detectable trend was also determined for sampling three and seven storms.

Stormwater Sampling and Analysis

Stormwater discharges were sampled for an entire water year (October 1, 1997, to September 30, 1998) on the Santa Ana River at W. 5th Street in Santa Ana, California, the last gauging station before discharge to the ocean. Automated stormwater samplers were installed that logged flow continuously and water quality samples were collected when flow rose above baseline conditions (0-0.7 m^3/s). Samples were collected at 15-min intervals; sampling intervals occasionally were extended to 30 min or an hour during tailing storm flows on extremely large storms when flow and runoff concentrations were changing slowly.

Over 1,700 stormwater samples were collected and analyzed for TSS, representing 90% of the total storm volume discharged during the 1998 water year. The remaining 10% of unsampled volume was a result of stormwater flow lower than the pump intake (7%) and equipment malfunction or breakage (3%). All stormwater samples were stored under refrigeration and analyzed for TSS because they are widely viewed as an indicator of stormwater quality and are correlated with other stormwater quality constituents (Sansalone and Buchberger 1997, Thomson *et al.* 1997). The TSS were analyzed by filtering a 10 to 100 mL aliquot of stormwater through a tarred 1.2 µm (micron) Whatman GF/C filter. The filters plus solids were dried at 60° C for 24 h, cooled, and weighed.

Sampling Designs and Estimators

Three designs for sampling within storms and five designs for sampling among storms were simulated and three estimators of mass emissions and concentration were used for each design. These separate, but dependent, issues were addressed in two steps. First, designs and estimators for use within storms were assessed in unison. Second, sample designs and estimators for annual mass emissions were addressed in unison using the optimal within-storm design and estimator.

The three within-storm sample designs were flow interval, time-interval, and simple random sampling. Flow-interval samples were taken at regular volume intervals (in practice, volume would be predicted beforehand based upon standard hydrologic principles or based upon historical data). This method provided samples that were evenly distributed with respect to volume during the storm; as flows and the volume discharged increased, so did the sample pacing. We considered designs of this type using sample sizes of 4, 8, and 12. Time-interval samples, representing the second design, were taken every 15 min for the first hour of the storm and one per hour thereafter, up to 96 h. This design had a random sample size, determined by the length of the storm. Sampling over time ensured that the samples were taken over the whole time range of the storm. The third design was a simple random sample of sizes 3, 4, 8, 12, and 42. Sample sizes of 4, 8, and 12 flow-paced samples and 42 time-interval sample sizes were chosen to compare designs based upon currently used compositing strategies and mean storm duration (Schiff 1997).

Three estimators of within-storm TSS mass emissions were compared (Table 1). The first estimator was the mean TSS from a storm multiplied by the total storm volume. The second estimator, volume weighted, was the product of the TSS and flow for each sample divided by the sum of sample flows and then multiplied by the total storm volume. This estimator adjusted TSS in a sample by the flow of the river during the sample. The third estimator was a ratio estimator (Cochran 1977). The product of TSS from samples and the ratio of total storm over sample volume was multiplied by total storm volume. This estimator assumed a positive relationship between TSS and volume.

Five designs were considered for sampling among storms: (1) stratified by size of storm (small, medium, and large); (2) stratified by season (early, mid-season, and late); (3) simple random sample from all storms; (4) simple random sample from only medium and large storms; and (5) sampling the first storms of the season up to a specified sample size (first *m*). Size strata were

TABLE 1. Within-storm estimators of TSS mass emissions.

Mean

$$\hat{y}_{j,1} = \left(\sum_{i=1}^{n} \frac{TSS_i}{n} \right) \left(\sum_{i=1}^{N_j} V_i \right)$$

Volume-weighted

$$\hat{y}_{j,2} = \left(\frac{\sum_{i=1}^{n} TSS_i \times V_i}{\sum_{i=1}^{n} V_i} \right) \left(\sum_{i=1}^{N_j} V_i \right)$$

Ratio

$$\hat{y}_{j,3} = \left(\frac{\sum_{i=1}^{n} TSS_i}{\sum_{i=1}^{n} V_i} \right) \left(\sum_{i=1}^{N_j} V_i \right)^2$$

$\hat{y}_j$ = Estimate of mass loading for storm j
n = Sample size
TSS_i = Total suspended solids in sample i
N_j = Number of possible samples in storm j
V_i = River volume between sample $i-1$ and sample i

created by grouping storms into small (vol < 1 x 10^9 l), medium (1 x 10^9 l < vol < 20 x 10^9 l), and large (vol > 20 x 10^9 l) categories. Seasonal stratification was accomplished by specifying early (September-January), mid (January-February), and late (February-May) season storms. Since stratified sampling requires allocation of the sample sizes among strata, all possible allocations were made for each sample size from 3 to 17 (the number of storms in the 1998 water year). Simple random sampling was achieved by Monte Carlo sampling of 1,000 from all possible samples for each sample size from 3 to 17. Simple random sampling of medium and large storms was accomplished using the same methodology for each sample of sizes 3 to 13, the number of medium and large storms.

Three estimators of annual mass emissions and concentration were considered: median, ratio, and regular (Table 2). The median estimator was the product of the median storm TSS concentration and the total volume of storms for the season. The ratio estimator was the product of the sum of TSS over all sampled storms and the ratio of total season and sample volume times the total season volume. This estimator assumed that TSS per storm was positively related to volume per storm. The regular estimator is the product of the total of TSS for the sampled storms and the ratio of the number of storms in the season to the number of storms sampled. To estimate annual TSS concentration, $\overline{Y}$ instead of concentration, we simply divided $\hat{Y}$ by the population multiplier; $\sum_{j=1}^{M} V_j$ for the median and ratio estimators and for the regular estimator.

Comparison of Designs and Estimators

The sampling designs and estimators were compared with respect to bias and precision. Bias was calculated as the average percentage of difference between the expected estimate and the actual results. The expected estimate was the average of the Monte Carlo samples or all possible samples, depending upon the design. Three designs had no random component: (1) time-interval within-storm design; (2) flow-interval within-storm design; and (3) sampling the first storms among-storm design. In these cases, the expected value was calculated from the one possible realization. The estimate was averaged over the sample size since little difference was found among them. Precision for all design/estimator combinations for various sample sizes was reported as a standard error, calculated as the square root of the variance of all calculated estimates. For the non-random designs, the standard error was calculated as the standard deviation of the data divided by the square root of the sample size.

TABLE 2. Among-storm estimators of annual TSS mass emissions.

Median

$$\hat{Y}_{median} = Median\left(\overline{y}_j\right) \times \sum_{j=1}^{M} V_j$$

Ratio

$$\hat{Y}_{ratio} = \left(\frac{\sum_{j=1}^{m} \hat{y}_j}{\sum_{j=1}^{m} V_j} \right) \left(\sum_{j=1}^{M} V_j \right)$$

Regular

$$\hat{Y}_{regular} = \frac{M}{m} \sum_{j=1}^{m} \hat{y}_j$$

$\hat{Y}$ = Estimate of annual mass emissions
$\overline{y}_j$ = Estimate of concentration for storm j
$\hat{y}_j$ = Estimate of mass emissions for storm j
m = Number of storms sampled
M = Number of storms
V_j = Volume of storm j

The appropriate annual sample size was assessed by comparing the confidence interval width for estimates of annual mass emissions and concentration. A 95% confidence interval was used based upon the optimal strategy.

The amount of detectable trend in annual mass emissions and concentration was calculated using the optimal strategy for sample sizes 3 and 7. A 90% confidence interval and 80% power were used. A linear trend was assumed based upon a regression setup

(Gerrodette 1987). This approach also assumed that the variability observed during the study is consistent from year to year. The assumption appears warranted, or at least conservative, as this year was an El Niño year that generated both typical and atypical storm patterns that varied tremendously in size and duration relative to historical rainfall patterns in the region.

RESULTS

The flow-interval sampling design with 12 samples provided the least bias of storm mass emissions (Table 3). Smaller sample sizes resulted in larger bias, as did both time-interval and simple random sampling within storms. No design consistently had the smallest standard error, but the standard error decreased as sample size increased.

The volume-weighted estimator was the best overall estimator of storm mass emissions (Table 4). It generally attained smaller bias than either the mean estimator or ratio estimator. The volume-weighted estimator also achieved the smallest standard error. Flow-interval sampling with the volume-weighted estimator estimated storm mass emissions that were too high approximately 65% of the time, whereas the simple random sample with median estimator and time-interval sample with volume-weighted estimator estimated mass emissions that were too low at least 65% of the time (Table 4). As a compromise, the flow-interval design with 12 samples and the volume-weighted estimator were used to characterize storms in the among-storm comparisons to achieve minimum bias with maximum precision.

The simple random sample of all storms or of medium and large storms resulted in the least bias in estimating annual TSS concentration (Table 5), but all designs attained similar bias in estimating annual TSS mass emissions (Table 6). Stratifying the results by season and first storms design resulted in the largest amount of bias in estimating annual TSS concentration. The simple random sample of medium and large storms provided the smallest standard error for concentration (Figure 1) and nearly the smallest standard error for mass emissions (Figure 2). Many design/estimator combinations are not included in these figures since their standard errors ranged from hundreds to thousands of times larger.

No estimator consistently gave the lowest bias in estimating annual TSS concentration (Table 5), but the ratio estimator was least biased, except for the first *m* storms design, for estimating annual TSS mass emissions (Table 6). The ratio estimator with eight or fewer storms sampled provided the smallest standard error for concentration (Figure 1). For mass emissions, the ratio estimator provided the smallest standard error (Figure 2). The regular estimator generally overestimated annual mass emissions by a larger number than did the ratio estimator (Tables 5 and 7).

The confidence interval width for annual concentration narrowed as sample size increased, but did not decrease proportionately for sampling more than seven storms (Figure 3). The confidence interval width for annual mass emissions decreased with increasing sample size, indicating no optimal sample size (Figure 4).

A 20% trend in mass emissions or concentration over five years was achieved by sampling three storms (Figure 5). The percent of detectable trend was reduced by 10 to 30% by increasing the sample sizes from 3 to 7, depending upon the number of years of interest. This relationship was the same for concentration and mass emissions.

DISCUSSION

We were able to assess the most efficient and effective monitoring design based upon a census of urban stormwater runoff for one wet season. The preferred within-storm design was a flow- or volume–paced strategy with the volume-weighted estimator. This estimator utilized the available volume or flow information more precisely than the ratio estimator. The preferred among-storm design was a simple random sample of medium- and large-sized storms using the ratio estimator for annual mass emissions or concentrations. Storm by storm, stronger positive correlation was observed between volume and TSS, thus making the ratio estimator more efficient. This

TABLE 3. Bias and precision of within-storm sampling designs and estimators of TSS mass emissions.

Design	Sample Size	Standard Error (MT)			Bias (%)		
		$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$
Flow Interval	4	4,295	2,958	372,019	3	-2	29
Flow Interval	8	3,066	1,450	191,902	3	5	29
Flow Interval	12	2,384	847	98,580	<1	2	24
Time Interval	42	969	135	14,687	-13	-7	7,570
Simple Random Sample	12	1,591	<1	5,207	-3	-100	-91

TABLE 4. TSS mass emissions estimates (in metric tons) for least biased combinations of within-storm designs (sample size) and estimators.

Storm Number	True TSS Loading	Flow Interval (12) and Volume-Weighted Estimator	Simple Random Sample (12) and Current Estimator	Time-Interval (42) and Volume-Weighted Estimator
3	946	835	1,156	625
5	18,436	19,093	18,137	16,973
6	320	338	296	291
7	5,705	6,179	4,957	4,935
9	21	19	22	22
10	5,278	6,174	5,068	4,667
11	31,846	32,602	35,084	28,433
12	17,967	19,239	13,081	17,967
13	867	947	573	917
14	91,628	93,422	107,015	75,446
15	943	883	900	913
16	81	81	84	74
17	3,090	2,504	2,975	3,097
18	1,556	1,636	1,460	1,531
19	623	593	655	583
21	2,107	2,289	1,718	2,031
22	9,499	10,939	9,399	8,361
Total	190,913	197,773	202,578	166,867

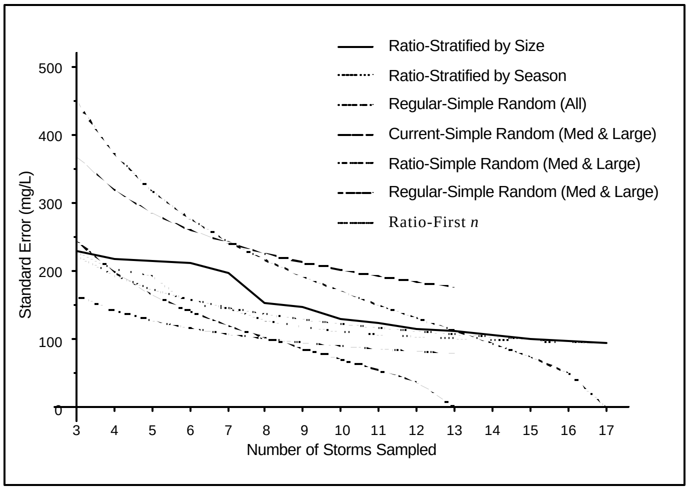
TABLE 5. Percent bias of TSS mass emission estimates for among-storm designs and estimators.

Design	$\hat{Y}_{median}$	$\hat{Y}_{ratio}$	$\hat{Y}_{regular}$
Stratified by Storm Size	-18	3	7
Stratified by Storm Season	-27	0	7
Simple Random Sample of All Storms	-27	4	6
Simple Random Sample of Medium & Large Storms	-18	4	6
Sample First *m* Storms	14	32	-1

TABLE 6. Percent bias of TSS concentration estimates for among storm designs and estimators.

Design	$\bar{Y}_{median}$	$\bar{Y}_{ratio}$	$\bar{Y}_{regular}$
Stratified by Storm Size	-6	13	8
Stratified by Storm Season	-17	32	29
Simple Random Sample of All Storms	-27	3	3
Simple Random Sample of Medium & Large Storms	-18	3	-7
Sample First *m* Storms	13	31	45

FIGURE 1. Standard error for estimates of annual TSS concentration for various estimator-sample design combinations.



overall combination provided the least bias relative to the actual mass emissions and smallest estimation of standard error.

The optimal strategy provided large improvements on current designs. A variety of designs have been used in southern California (Schiff 1997); one common approach is to use 4 to 12 flow-weighted samples within storms with a mean estimator and a median estimator for annual mass emissions and concentration. The selection of storms for sampling is generally subjective with preference given to large or early storms. This current design has little bias for within-storm estimates but has a standard error almost three times the optimal. The annual estimates of mass emissions and concentration from the median estimator, even with optimal within-storm design, have large negative bias and standard error that are two to three times as large as the optimal design. The correction to achieve more accurate and precise estimates is a simple matter of changing estimators and applying simple modifications to sample and storm selection methodologies.

Although these findings will improve the knowledge of stormwater discharges in an arid region, there are limits to our conclusions. The largest of these limits is our ability to extrapolate our findings to other watersheds. If the variance in water quality changes, or if the relationship between flow and water quality differs, then our conclusions for the Santa Ana River data may not apply. We

FIGURE 2. Standard error for estimates of annual TSS mass emissions for various estimator-sample design combinations.

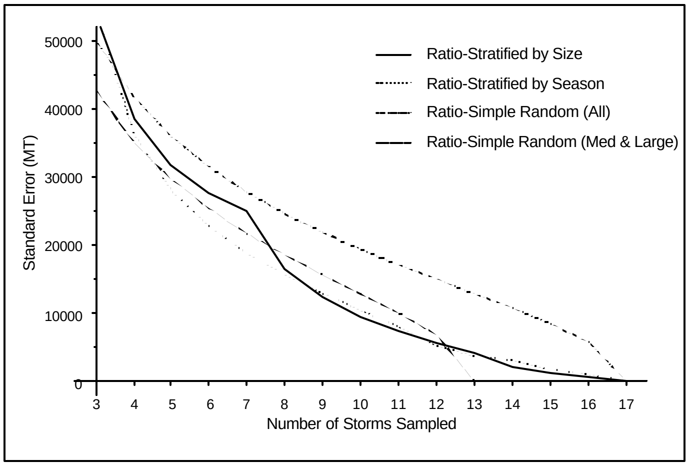


TABLE 7. Annual mass emissions estimates of TSS (in metric tons) for among-storm designs and estimators (number of storms, *m*=7).

Design/ estimator	Annual mass loading (ratio/ estimator)	Annual mass loading (regular/ estimator)
Actual Load= 190,913		
Stratified by Storm Size	197,329	204,432
Stratified by Storm Season	191,418	204,623
Simple Random Sample of All Storms	199,802	203,044
Simple Random Sample of Medium & Large Storms	198,523	202,980
Sample First *m* Storms	253,606	189,796

FIGURE 3. 95% confidence intervals for ratio estimate of annual TSS concentration from simple random sample of medium and large storms (mean = 820 mg/L).

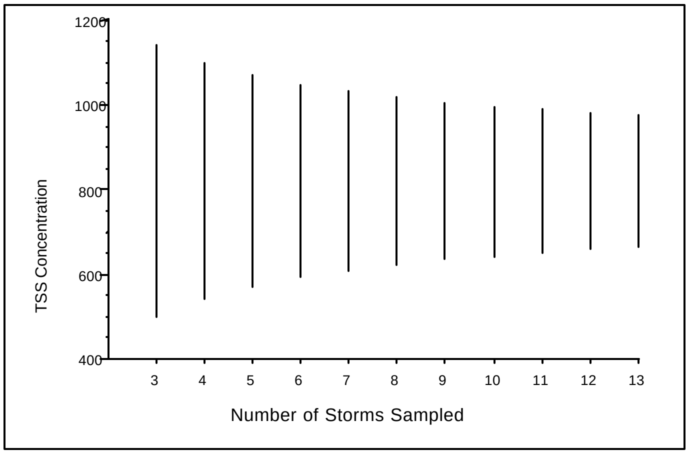


FIGURE 4. 95% confidence intervals for ratio estimate of annual TSS mass emissions from simple random sample of medium and large storms (mass emissions = 6.5 billion metric tons).

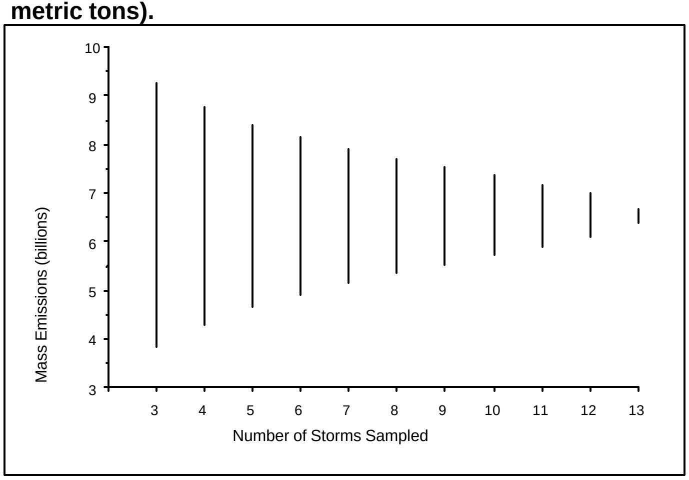
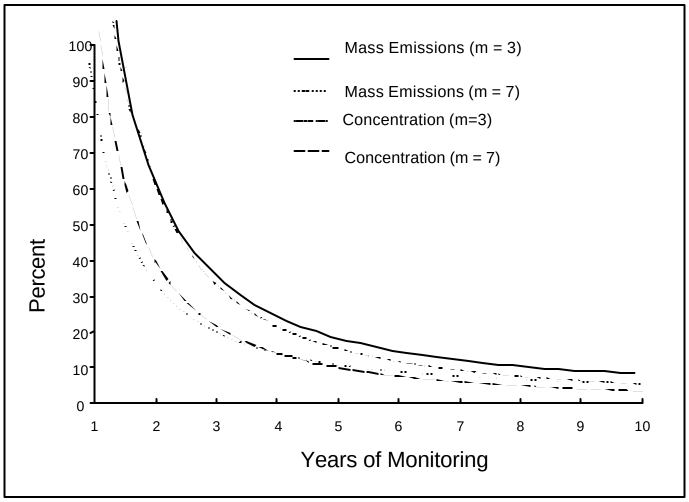


FIGURE 5. Percent detectable trend over years for annual mass emissions and concentration using the ratio estimator with simple random sample of medium and large storms.



suspect that the Santa Ana River is similar to other large, urbanized channels in southern California, but is different from outlets in the rest of the country.

A second limitation is our ability to extrapolate to other years. The 1997/98 season was anomalously wet. Our findings were based upon the wider variability associated with the El Niño effect, so our estimates of standard error were larger also. Because few wet seasons are expected to introduce more variability, we have made conservative estimates of precision and trend detection and therefore suspect that our recommendations would still apply.

Another limitation to this study is the difficulty in extrapolating our conclusions to other constituents. In this study, we utilized TSS since it is an easily measured constituent with relatively small laboratory variability suited for exploring different designs. However, we did analyze trace metals, total organic carbon (TOC) and total nitrogen (TN) on a subset of samples and found these constituents to be highly correlated with TSS (Tiefenthaler *et al.*, 2001). Similarly, additional methodologies are available that measure sediment mass emissions and suspended sediments. The methodology utilized in this study was patterned after urban stormwater programs required by National Pollutant Discharge Elimination System Permit reporting and monitoring for municipal separate stormwater sewer systems established by state and federal regulatory agencies (e.g., the U.S. EPA). However, the U.S. Geological Survey (Porterfield 1972) utilizes very different sampling and analytical techniques to estimate fluvial sediment transport. The extent to which these methodologies could alter our assessments of optimal designs is unknown.

The inherent advantage of the volume-paced sampling method was the significant relationship that has been established between flow and TSS in previous studies. In fact, flow accounted for 40% of the variability in TSS concentrations in the Santa Ana River during the 1997/98 wet season (Tiefenthaler *et al., this volume*). Moreover, flow and TSS concentration are highly correlated in all 12 of the largest rivers and creeks in the SCB (Cross *et al.* 1992). Hence, sampling the medium and large storms of the year was a logical endpoint because mass emissions from this river are dominated by the larger episodic events. During 1997/98, more than 95% of the mass emissions from the Santa Ana River occurred during less than 5% of the year; this period was associated with the maximum flows of the wet season.

Trend detection in urban watersheds has been largely overlooked since the necessary data were not available until this present study. Relatively small changes (i.e., 10 to 20%) in both concentration and mass emissions over relatively small time scales (five years or less) can be detected if the monitoring designs are optimized.

LITERATURE CITED

Bay, S. and K. Schiff. 1997. Impacts of stormwater on the nearshore environment of Santa Monica Bay. pp. 105-118 *in*: S. Weisberg, C. Francisco, and D. Hallock (eds.), Southern California Coastal Water Research Project Annual Report 1995-1996. Westminster, CA.

Cochran, W.G. 1977. Sampling Techniques. John Wiley & Sons, Inc. New York, NY.

Cross, J., K. Schiff and H. Schaefer. 1992. Surface runoff to the Southern California Bight. pp. 19-28 *in*: J. Cross (ed). Southern California Coastal Water Research Project Annual Report 1989-1990. Southern California Coastal Water Research Project. Long Beach, CA.

Gerrodette, T. 1987. A power analysis for detecting trends. *Ecology* 68:1363-1372.

Haile, R.W., J.S. Witte, M. Gold, R. Cressey, C. McGee, R.C. Millikan, A. Glasser, N. Harawa, C. Ervin, P. Harmon, J. Harper, J. Dermand, J. Alamillo, K. Barrett, M. Nides and G.-Y. Wang. 1999. The health effects of swimming in ocean water contaminated by storm drain runoff. *Epidemiology* 10: 355-363.

Porterfield, G. 1972. Computation of fluvial sediment discharge. U.S. Geological Survey, Department of the Interior. U.S. Government Printing Office. Arlington, VA.

Sansalone, J.J. and S.G. Buchberger. 1997. Partitioning and first flush of metals in urban roadway storm water. *Journal of Environmental Engineering* 123:134.

Schiff, K. 1997. Review of existing stormwater monitoring programs for estimating bight-wide mass emissions from urban runoff. pp. 44-55 *in*: S. Weisberg, C. Francisco, and D. Hallock (eds.), Southern California Coastal Water Research Project Annual Report 1995-1996. Westminster, CA.

Schiff, K and L. Tiefenthaler. 2001. Anthropogenic versus natural mass emissions from an urban watershed. pp. 63-70 *in:* This annual report. S. Weisberg (ed.), Southern California Coastal Water Research Project Annual Report 1999-2000. Southern California Coastal Water Research Project. Westminster, CA.

Tiefenthaler, L, K. Schiff and M. Leecaster. 2001. Temporal variability patterns of stormwater concentrations in urban stormwater runoff. pp. 52-62 *in:* This annual report. S. Weisberg (ed.), Southern California Coastal Water Research Project Annual Report 1999-2000. Southern California Coastal Water Research Project. Westminster, CA.

Thomson, N.R., E.A. McBean, W. Snodgrass, and I.B. Monstrenko. 1997. Highway stormwater runoff quality: Development of surrogate parameter relationships. *Water, Air, Soil Pollution* 94:307.

ACKNOWLEDGEMENTS

The authors are extremely grateful for the field and laboratory personnel that assisted with this project: Dario Diehl, David Tsukada, Jeff Brown, Shelly Moore, and Andrew Jirik. A special note of appreciation is given to Harold Stubbs.