WATER RESEARCH 41 (2007) 4186–4196

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/watres

Design of stormwater monitoring programs

Haejin Lee[a], Xavier Swamikannu[b], Dan Radulescu[b], Seung-jai Kim[c], Michael K. Stenstrom[d,*]

[a]United States Department of Agriculture, Natural Resources Conservation Service, El Centro Office, 177 N. Imperial Avenue, El Centro, CA 92243, USA
[b]California Environmental Protection Agency, Los Angeles Regional Water Quality Control Board, 320 West 4th St. Suite 200, Los Angeles, CA 90013-2343, USA
[c]Department of Environmental Engineering, Chonnam National University, 300 Yongbong-dong, Buk-gu, Gwangju 500-757, Korea
[d]Department of Civil and Environmental Engineering, University of California, 5714 Boelter Hall, Los Angeles, Los Angeles, CA 90095, USA

ARTICLE INFO

Article history:
Received 18 December 2005
Received in revised form
4 January 2007
Accepted 10 May 2007
Available online 18 May 2007

Keywords:
First flush
Industrial general permit
Monitoring
Municipal permit
Stormwater

ABSTRACT

Stormwater runoff is now the leading source of water pollution in the United States, and stormwater monitoring programs have only recently been developed. This paper evaluates several stormwater monitoring programs to identify ways of increasing the likelihood of identification of high-risk dischargers and increasing data reliability for assisting in the development of total maximum daily loads. No relationship was found between various types of industrial activity or landuse and water quality data. Stormwater data collected with grab samples has much greater pollutant concentration variability than in potable water or wastewater monitoring programs. Industrial land use is an important source of metals. For grab samples, sampling time during the storm event will affect results. Data from California, which has distinct dry periods, showed a seasonal first flush, whereas data from areas with more uniform rainfall throughout the year did not show a seasonal first flush. Selecting a subset of sites from each monitored category using a flow-weighted composite sampler is an alternative strategy, and may result in lower overall cost with improved accuracy and variability in mass emissions, but may not be less successful in identifying high-risk dischargers.

1. Introduction

The completion of wastewater treatment plants mandated by the Clean Water Act has reduced pollution from point sources to the waters of the United States. Parallel trends are occurring in other developed countries and will occur in newly developing countries over time. As a result, non-point source pollution such as stormwater runoff is now the major contributor to pollution of receiving waters. The problem of stormwater pollution is growing worse because of continuing development, which results in increased impervious surface area. In response to this growth, regulatory agencies are requiring stormwater monitoring programs implemented through the National Pollutant Discharge Elimination System (NPDES) to quantify and eventually reduce stormwater pollution.

The overall goals of stormwater monitoring programs include the identification of high-risk dischargers, the development of total maximum daily loads (TMDLs) and ultimately the reduction of stormwater pollution. Most stormwater monitoring programs are relatively new and evaluations of their usefulness for satisfying these goals are only now possible (Duke et al., 1998; Lee and Stenstrom, 2005; Pitt et al., 2003; Olding et al., 2004; Turner and Boner, 2004). We evaluated 9 years of data generated by industrial facilities

*Corresponding author. Tel.: +1 310 825 1408; fax: +1 310 206 5476.
E-mail address: stenstro@seas.ucla.edu (M.K. Stenstrom).
0043-1354/$ - see front matter
doi:10.1016/j.watres.2007.05.016

in Los Angeles County covered under the statewide Industrial StormWater General Permit (ISGP). The results of the study suggest that parts of the current industrial stormwater monitoring programs will not be helpful to identify high-risk dischargers, nor will they be useful in estimating mass emissions for developing TMDLs. The design and requirements of the monitoring programs do not produce data with sufficient precision for decision making.

We also evaluated a number of stormwater monitoring programs across the United States to determine their usefulness in achieving the dual goals, and we make recommendations for improvements in monitoring programs. Our analysis included three Industrial StormWater General Permit programs (ISGPs), Municipal StormWater Permit programs from 17 states and Los Angeles County, and UCLA's first flush highway runoff characterization study sponsored by California Department of Transportation (Caltrans). Our results should be helpful to planners and regulators to interpret existing datasets and programs, as well as provide recommendations for improving future programs.

2. Materials and methods

2.1. Data sources

Six stormwater monitoring programs were evaluated and are summarized in Table 1. Three of the monitoring programs were associated with industrial permits, two with monitoring of municipal separate storm sewer systems, and the last was associated with a research project to quantify discharges from highways.

The ISGP had the most observations. ISGP programs require facilities that discharge stormwater associated with industrial activities to apply and obtain coverage under the ISGP. Industries are categorized by Standard Industrial Classification (SIC) codes. These monitoring programs serve as a model for other programs in the United States and can be useful in developing monitoring programs for other countries. In our previous study, we analyzed data from the ISGP in Los Angeles County from three wet seasons (1998–2001). In this study, we extended the evaluations to nine wet seasons (1992–2001), and we also analyzed eight wet seasons (1993–2001) of data from a similar ISGP in Sacramento County. Since industrial monitoring programs generally vary by state, 9 years (1995–2003) of stormwater data from the State of Connecticut were also analyzed.

The StormWater Program for Municipal Separate Storm Sewer Systems (MS4) is a nationwide program and is designed to monitor and reduce sediment and pollution that enters surface waters from separate storm sewer systems to the maximum extent practicable. Under US EPA sponsorship, Pitt et al. (2003) compiled and evaluated stormwater data from a representative number of NPDES MS4 stormwater permittees. Over 10 years of data from more than 200 municipalities throughout the United States were assembled. The areas were primarily located in the southern, Atlantic, central and western parts of the United States. Only data from well-described stormwater outfall locations were used in the database. The Los Angeles County MS4 data, which were not included by Pitt et al. (2003), were analyzed and added as a separate program, as shown in Table 1.

The Department of Civil and Environmental Engineering at UCLA has characterized stormwater runoff from three highway sites for Caltrans since 1999 (Kim et al., 2005; Khan et al., 2006). The study was conducted to assess runoff water quality and quantity from California freeways, with particular emphasis on characterizing runoff during the early stage of storm events or the first flush. The study used grab and flow-weighted composite samples and measured a large range of parameters.

Regulators and others need to understand that decision making using stormwater monitoring results collected under the current regulatory scheme is limited due to the data's high variability. The variability is associated with natural variability of stormwater, but is compounded by experimental error, which is introduced through limited or poor sampling and analysis techniques. To illustrate this difference, we analyzed the coefficient of variation (CV) of other monitoring activities and compared it to storm-monitoring variability. For this analysis, two extra datasets that are not shown in Table 1 were used for comparison purpose. The two datasets are from two typical water and wastewater treatment plants in Southern California. The data from these plants were collected over an entire year and used composite samples and certified laboratories for analysis.

2.2. Methods

These datasets and others were all analyzed used Systat 10.2 (Systat Software, Inc., Point Richmond, CA), which is a well-known program for performing a range of statistical analyses. Means, standard deviations, coefficients of variations and other simple statistics were performed using Systat procedure Descriptive Statistics. Sysat's General Linear Models (GLM) Routine was used to perform analysis of variance (ANOVA), and data for Fig. 7 was generated using Systat's T-test procedure.

Sampling methods, either grab or composite, are noted in the description of the datasets and have large implications for the variability and utility of the datasets. Grab samples are discrete samples taken within a short period of time, usually less than 15 min, and can be collected any time during the storm event. The time of sampling, either early or late in the runoff, can have important impacts and these impacts are also discussed.

The datasets were collected directly form the agencies responsible for the monitoring programs in machine-readable formats. There was no opportunity to introduce coding or transcription errors. The datasets were transmitted to the regulatory agencies in the same way as we received them. The only quality assurance procedures applied to the datasets were those implemented by the agencies themselves.

Previous studies, even the original Nationwide Urban Runoff Program (NURP) study (US EPA, 1983) and previous work from our laboratory (Stenstrom et al., 1984; Fam et al., 1987; Lau and Stenstrom, 2003), identified relationships between water quality and land use in most cases. We expected to see differences in water quality associated with different types of landuse or industrial activity. In our previous work (Ha and

Table 1 – Summary of major monitoring datasets for this study

Monitoring program	Monitoring area	Primary land use	Period	Sampling type	Monitored parameters[a]	Number of observations
Industrial General Permit	County of Los Angeles, CA	Industrial	1992–2001	Grab	Mandatory parameters are pH, SC, TSS, and O&G or TOC Certain facilities must analyze metals	24,852[b]
Industrial General Permit	State of Connecticut	Industrial	1995–2003	Grab	Mandatory parameters are pH, COD, O&G, TSS, Cu, Pb, Zn, TP, TKN, NO_3, and LC_{50}	9589
Industrial General Permit	County of Sacramento, CA	Industrial	1993–2001	Grab	Mandatory parameters are pH, SC, TSS, and O&G or TOC Certain facilities must analyze metals	810[b]
Municipal Permit	17 states in the US	Various land uses	1991–2002	Flow-weighted composite Several pollutants such as bacterial indicators and O&G were collected by grab sample	Nutrients, TSS, heavy metals, and organic toxicants, vary by state	3765[c]
Municipal Permit	County of Los Angeles, CA	Various land uses	1996–2001	Flow-weighted composite Several pollutants such as bacterial indicators and O&G were collected by grab sample	Nutrients, heavy metals, and organic toxicants such as pesticides and herbicides	184
First Flush Highway Runoff Characterization	405 and 101 freeway near UCLA, CA	Highway	2000–2002	Grab and flow-weighted composite	TSS, VSS, bacterial indicators, nutrients, O&G, COD, DOC, metals, hardness, and alkalinity	351[d]

[a] Specific conductance (SC), total suspended solids (TSS), volatile suspended solids (VSS), oil and grease (O&G), total organic carbon (TOC), dissolved organic carbon (DOC), chemical oxygen demand (COD), copper (Cu), lead (Pb), zinc (Zn), aluminum (Al), iron (Fe), total phosphorus (TP), total Kjeldahl nitrogen (TKN), and nitrate (NO_3).
[b] The number of observations by parameter is shown in Table 2.
[c] See Pitt (2004) for the number of observations for each parameter.
[d] Sample by grab and flow-weighted composite sample since 1999. However, we used only grab samples and EMCs calculated from grab samples in this paper.

Stenstrom, 2003), we successfully applied neural networks (NNs) to identify relationships between water quality data and various types of landuse (commercial, residential, industrial, transportation, and vacant) using Los Angeles County MS4 data. In this study, we used the same techniques to analyze the ISGP data from the County of Los Angeles, the County of Sacramento, and the State of Connecticut to identify the various industrial activities as a function of stormwater quality data. A multi-layer perception neural model, the most common supervised NN, was used to differentiate the various types of industrial activity using Neural Connection 2.1 (SPSS, Inc. and Recognition, Inc., Chicago, IL).

Industrial and transportation landuses are generally known as greater sources of heavy metals (generally six metals: cadmium, copper, chromium, lead, nickel, and zinc) than other land uses such as residential land use. We evaluated the concentrations of metals in runoff from various landuse or industrial categories to determine if landuse or source could be associated with stormwater concentrations. We also compared concentrations for each monitoring program, when possible, to determine how many were outside the US EPA's stormwater quality benchmarks.

The timing of a grab sample is important since the first flush associated with many stormwaters creates a higher concentration at the beginning of runoff. Therefore, the sampling time during a storm event is important to avoid the bias of the first flush and better characterize the mass emission of the event. For example, Khan et al. (2006) showed that oil and grease concentrations in samples from highways in the first 15 min of runoff were several times higher than the event mean concentrations (EMC). Samples early in a storm event should be collected if the peak or maximum concentrations are desired.

The industrial sites are generally small, impervious watersheds and should experience a first flush. In so far as possible, we have identified the sampling time and noted the implications of sampling time on monitoring results.

The decision of which storm events to sample is also problematic. Climatic conditions, such as the existence of long dry periods, may greatly impact pollutant emissions from urban stormwater discharges. Most western parts of the United States, including Los Angeles and Sacramento are dry from May to September. This rainfall pattern creates a long period for pollutant build-up and, therefore, the initial storm of the wet season may have higher pollutant concentrations than in later events (Lee et al., 2004). This phenomenon is called a "seasonal first flush," and can be illustrated using ISGP data from Los Angeles. These same climatic conditions can be found in southern Europe, areas in South America, Africa, and Australia.

The number of storms to be sampled during each monitoring season is also an important parameter. Our analysis does not address this decision, but we note that only one or two storms each year is not likely to be representative. Leecaster et al. (2002), in examining several monitoring programs in California, recommended sampling seven storms per year (approximately 50% of the storms in Southern California in a typical year) to obtain small confidence intervals. Increasing the number of storms or samples will create a burden to all permittees, who will understandably question the benefits. We suggest an alternative strategy to reduce the burden of increased sampling requirements.

3. Results and discussion

Table 2 shows the basic statistics and US EPA benchmark levels for the Los Angeles, Sacramento, and State of Connecticut ISGP programs. The number of samples, mean, and CV are shown. These results will be used in several of the following sections.

3.1. Sampling method

Fig. 1 shows the CV (equal to the standard deviation divided by the mean) for various routinely monitored water quality parameters for water, wastewater and stormwater management. The leftmost bars show the CVs for influent wastewater quality parameters from a large west-coast wastewater treatment plant. The CVs range from 0.09 to 0.2. The next set of bars show the CVs for raw water quality parameters for a large drinking water treatment plant. The CVs range from 0.07 to 0.3. The next group of bars shows the CVs from the Municipal StormWater Permits (MS4), which range from 1.1 to 3.3 (Pitt, 2004). The final group of bars shows the CVs from the ISGP data for Los Angeles County that are also tabulated in Table 2. The ISGP data for Los Angeles County have the highest CV among the various water data, and ranged from 2.8 to 14. The CVs from the ISGP data for the County of Sacramento and the State of Connecticut are shown in Table 2 are also high when compared to the water and wastewater monitoring results.

The variability in the stormwater data is several times the mean value. With such high variability, it is virtually impossible to make statistical inferences. Regulators and users of stormwater monitoring data need to recalibrate their expectations from stormwater monitoring programs. The challenge in developing better monitoring programs is to reduce the variability due to sampling errors and artifacts, while retaining the intrinsic variability.

The sampling method is a major source of great variation. For ISGP monitoring, grab samples are allowed, whereas flow-weighted composite samples were collected in the

Table 2 – Basic statistics for three industrial general permit programs

Parameter	Unit	Benchmark level[a]	Los Angeles County			Sacramento County			Connecticut State		
			Sample no.	Mean	CV	Sample no.	Mean	CV	Sample no.	Mean	CV
pH	pH unit	6.0–9.0	24,851	7.01	0.95	857	7.16	0.17	9617	6.32	0.20
TSS	mg/L	100	24,144	376	11.8	769	185	2.86	9617	124.	6.59
SC	μmos/cm	200	23,585	562	8.13	846	204	2.27			
O&G	mg/L	15	18 637	16.6	14.3	286	11.3	1.61			
TOC	mg/L	110	9714	50.1	5.23	399	31.4	2.12			
BOD	mg/L	30	728	165.	10.7						
COD	mg/L	120	1834	271.	2.77	50	154	1.58	9606	80.7	3.44
Al	mg/L	0.75	1618	10.1	12.2	46	3.44	1.66			
Cu	mg/L	0.0636	3354	1.01	16.5	83	0.18	2.31	9596	0.13	7.56
Fe	mg/L	1	1844	25.5	6.39	82	7.49	2.03			
Pb	mg/L	0.0816	3525	2.96	14.1	78	4.48	3.82	9563	0.06	8.37
Ni	mg/L	1.42	1858	0.63	16.0						
Zn	mg/L	0.117	5163	4.96	13.9	141	2.23	7.59	9614	0.51	7.79
P	mg/L								9606	0.45	4.30
TKN	mg/L								9608	2.50	3.11
NO_3	mg/L								9613	1.19	2.73
24 h LC_{50}	%								9628	0.82	0.38
48 h LC_{50}	%								9628	0.75	0.45

[a] Benchmark level suggested by US EPA.

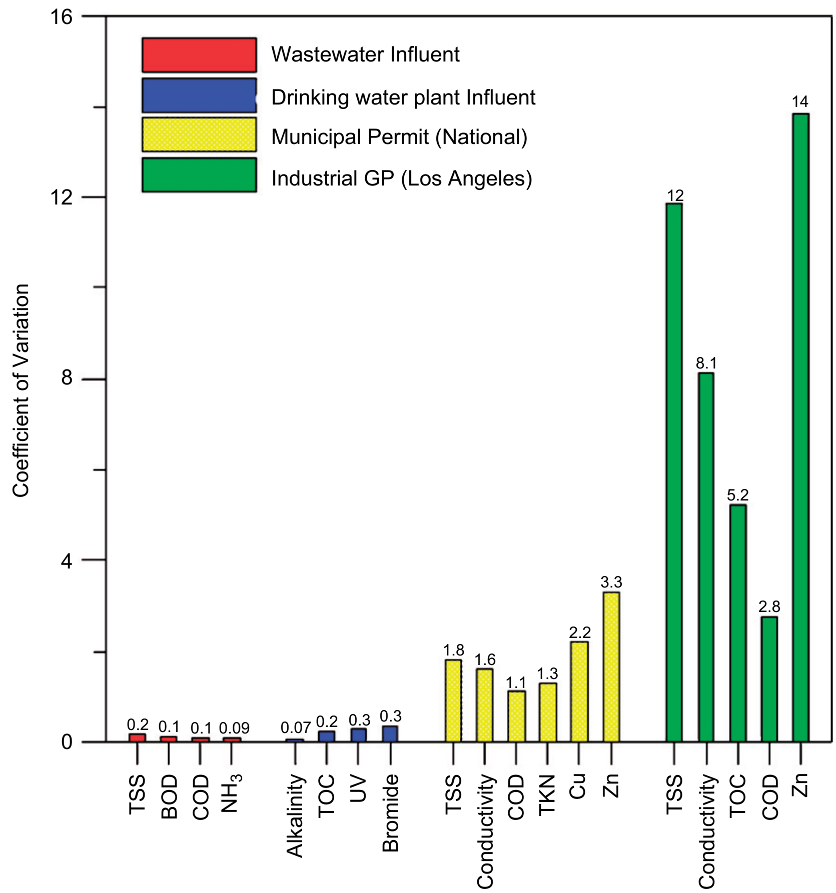

Fig. 1 – Coefficient of variation in various water sampling programs for a large wastewater treatment plant, a drinking water treatment plant, and two stormwater monitoring programs, illustrating the different amounts of variability.

Municipal StormWater Permit program and water and wastewater treatment plant monitoring programs. The use of grab samples must increase the variability and CV. It is universally recognized that collecting flow-weighted composite samples is better for stormwater monitoring. The result from a flow-weighted composite sample is often called an EMC, which is not only more representative than a grab sample, but can also be used to estimate pollutant loading since the product of EMC and total runoff volume is the pollutant load (Sansalone and Buchberger, 1997; Lee et al., 2002).

To illustrate the increased variability of grab samples over composite samples, Fig. 2 is provided, which shows sampling results from three storm events collected in a recently completed highway runoff study (Stenstrom and Kayhanian, 2005). The results are reported for three different highway runoff sites for three different storm events, and are representative of approximately 70 storm events monitored during the study. Values for hardness, chemical oxygen demand (COD), dissolved organic carbon (DOC), oil and grease (O&G), and ammonia (as N) are reported. Fig. 2 shows the value of each grab sample, and compares it to the composite sample, collected by a flow proportional automatic sampler, which is denoted as "E" on the graphs (for EMC). No composite is reported for O&G, which is usually not measured with automated samplers. It is easy to observe that the various grab samples can be 10 times greater or smaller than the mean or EMC. Hence, the use of a single grab sample to estimate mass emission rates or environmental impacts have large error. Averaging 12 grab samples (horizontal line in Fig. 2) is a much better estimate of the EMC, but still can be in error.

Unfortunately, collecting flow-weighted composite samples is more difficult and expensive. The location and installation of the sampler generally requires engineering and construction of sampling facilities, and training to operate the samplers. There are approximately 3000 industrial permittees in Los Angeles County alone, as compared to less than 50 monitoring programs for water and wastewater treatment plants. Requiring composite sampling represents a large financial burden. In addition, several water quality parameters, such as O&G, toxicity, and indicator bacteria are not easily measured by automatic composite samplers.

3.2. *Relationship between landuse activity and water quality data*

In a previous study, a multi-layer perception neural model (Nerual Connection 2.1, SPSS Inc., Chicago, IL) was trained with the basic water quality parameters for stormwater data collected from specific landuses using composite samplers (Lee and Stenstrom, 2005). The approach was applied to the Los Angeles and Sacramento County ISGP data shown in Table 2, and expanded to use three supervised, feed-forward

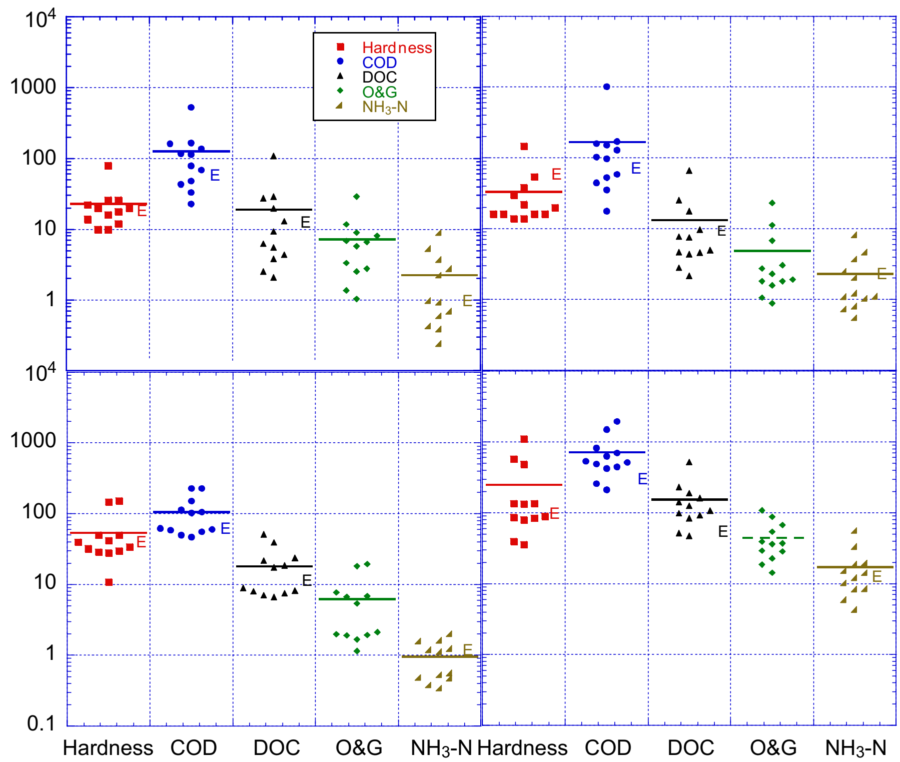

Fig. 2 – Scatter diagram of the values of grab samples for hardness (mg/l $CaCO_3$), chemical oxygen demand (COD), dissolved organic carbon (DOC), oil and grease (O&G), and ammonia nitrogen (NH_3-N), compared to the composite samples (E) collected with automatic flow-weighted samplers and the mean of the grab samples (horizontal line). Composite samples not collected for oil and grease. Upper left panel: Site 3, February 11, 2005; upper right panel: Site 2, October 26, 2004; lower right panel: Site 2, October 16, 2004; lower left panel: Site 1, October 26, 2004.

NNs (multi-layer perceprton, radial basis function, and Bayesian network). Because of the small number of observations, it was not possible to train using metals data. For Connecticut's data, both basic water quality parameters and metals data were sufficient to be used as input parameters. For output parameters, eight major industrial categories (based on SIC code) were selected, based on their prevalence in case number—food and kindred products (20); chemical and allied products (28); primary metal industries (33); fabricated metal products (34); transportation equipment (37); motor freight transportation and warehousing (42); electric, gas and sanitary services (49); and whole trade-durable goods (50). Outliers in the dataset were defined as greater than 1.5 times the interquartile range plus the 75% data or less than 25% data minus 1.5 times the interquartile range, and were removed from the analysis. This procedure is consistent with techniques used and documented in the Systat manuals. The neural models were extensively trained using various architectures; however, the performance of all models was very poor. Only 20% of data case are correctly classified which is not significant. These results confirmed that either no relationship exists between the industrial categories based on SIC code and the stormwater runoff quality, or that sampling errors are so large that they obscure any relationship.

The overall conclusion of this exercise is that the current monitoring programs are unable to discern a relationship between industrial activity and stormwater quality. The causes are probably a combination of imprecise nature of SIC codes as well as the errors introduced by poor monitoring technique, such as the use of grab samples and non-standard timing for sample collection.

3.3. *Sampling parameters*

Fig. 3 shows metal concentrations as a function of landuse (highway data from Han et al., 2006; other landuse calculated from Pitt's (2004) dataset. Industrial landuse has the highest median concentrations for cadmium, chromium, lead, and nickel, but copper and zinc are the highest from highways. Similar results were found from Los Angeles County's Municipal StormWater Permit program, which were not included in Pitt's dataset. Industrial landuse had the highest concentrations of aluminum, nickel, and zinc, but copper was highest in transportation landuse (not shown in this paper). The results confirm that industrial landuses are greater sources of heavy metals than other landuses. In addition, the concentration of metals more frequently exceeds the US EPA's stormwater benchmark than for basic water quality parameters (Fig. 4). For example, zinc concentrations exceeded the benchmark concentration approximately 90% of samples from the ISGP in Los Angeles County.

Including metals in industrial stormwater monitoring programs should be a high priority, but their presence varies among states and industrial categories. It is mandatory in

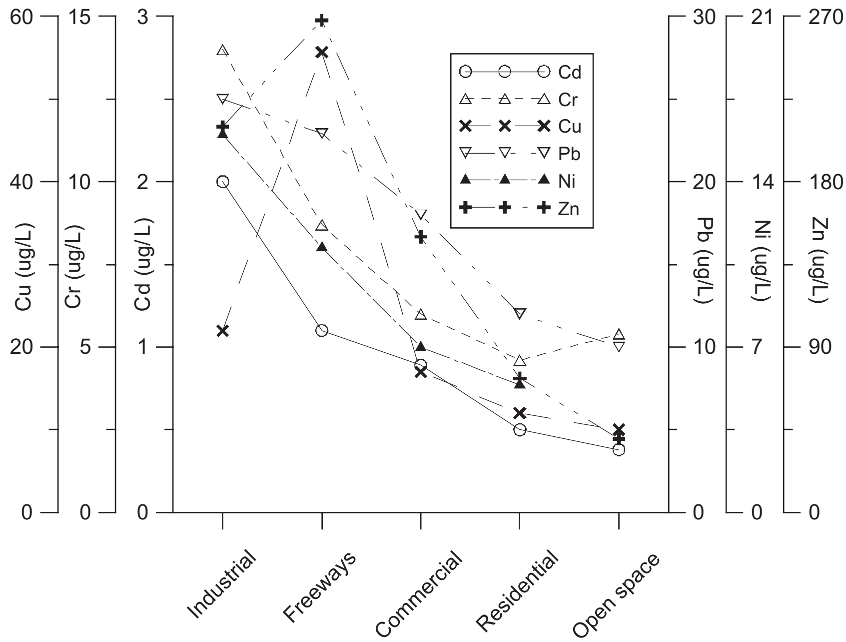

Fig. 3 – Median concentrations of metals in stormwater runoff from industrial, freeway, commercial, residential, and open land uses.

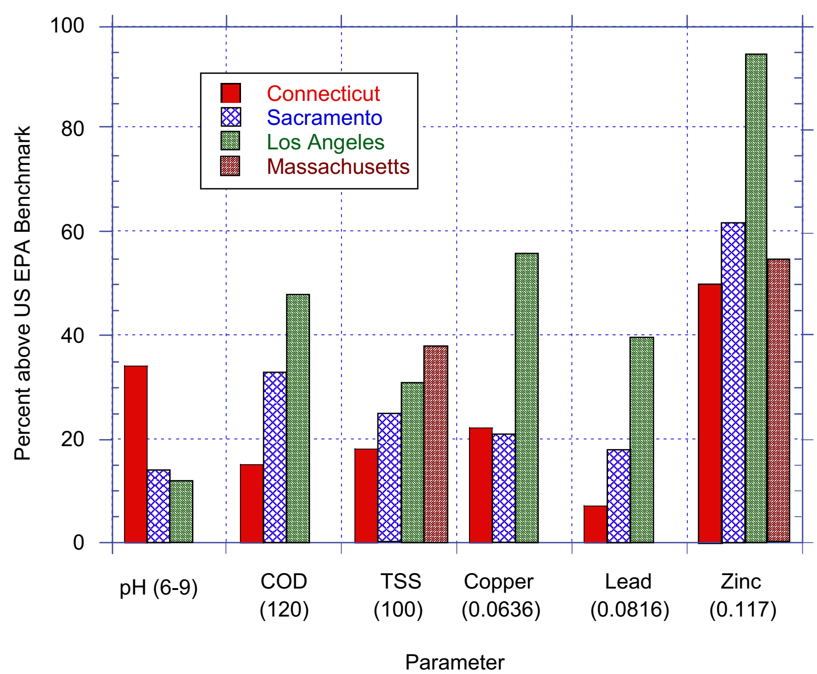

Fig. 4 – Percent of observations greater than US EPA Benchmark concentrations from four industrial permits (numbers below parameter value indicate US EPA benchmark value in mg/L, except for pH). Values from Massachusetts not available for all parameters.

Connecticut for all industrial categories, whereas only certain industrial facilities are required to analyze for metals in California. There may be logic for including or excluding metals in specific industrial permits, but the current monitoring results showed no remarkable differences in monitored pollutant parameters among industrial categories. The source of metals from industrial landuse is not only from industrial activity, but also may be from building materials, such as roof material and siding material of the industrial facilities (Davis et al., 2001). In anticipation of controlling metals for TMDLs and other purposes, any future permit should require industrial stormwater monitoring for metals such as cadmium, copper, chromium, lead, nickel and zinc, unless they can demonstrate that these metals are not present in their stormwater runoff.

3.4. Sampling time in a storm event

In a the recently completed highway stormwater monitoring project (Stenstrom and Kayhanian, 2005), grab samples were collected every 15 min during the first hour of the storm and additional grab samples were collected each hour for up to 8 h. Fig. 5 shows the impact of sampling time for TSS and total Zn from highway sites. The ratio of observed concentration to EMC is shown for more than 30 events for two wet seasons. If the grab samples were the same as EMC, the points would all appear along the horizontal line at 1.0. In general, the ratio of the observed concentration to EMC is much higher than 1.0 at the beginning of the storm and declines as the storm progresses. It is obvious that collecting a sample in the early part of the storm overestimates the EMC and the total load. Khan et al. (2006) has examined this effect for sampling O&G and concluded that collecting a grab sample 2–3 h into a typical storm more closely approximates the EMC than sampling earlier or later in the storm. If grab samples are to be used, the most appropriate time to sample should be investigated.

3.5. Sampling time during wet seasons

Fig. 6 shows the median concentrations of specific conductance, TOC, TSS, and Zn as a function of the normalized cumulative precipitation for ISGP data from Los Angeles County. The concentrations of all parameters were highest in the first few events, which is due to accumulation during the long dry summer, and decreased as the season progressed. A substantial concentration peak in the initial storm event suggests the presence of a seasonal first flush. If only the first storm is monitored, it will overestimate pollutant concentrations in later storms. However, areas with more uniform rainfall, such as Connecticut, did not shown a seasonal first flush. This presence of a seasonal first flush is a dilemma for monitoring programs, which attempt to identify high dischargers, as well as estimate central tendencies, such as annual emissions.

3.6. Sampling strategy

A routine response to inadequate data is to require more data collection. It is temping to assume that additional sampling will eventually overcome the variability of stormwater monitoring programs. This conclusion is generally false, because of the high variability. Fig. 7 shows the results of a simulating a two-tailed t-test to determine the number of observations required to detect differences in means between two datasets. An α value (significance) of 0.05 and a power of 0.8 were assumed. The graph shows how many samples are required to detect a statistically significant difference in means. The graph is plotted as a number of required samples as a function of relative difference in means, with the various lines corresponding to different coefficients of variation. The symbols indicate calculation points, with the numbers corresponding to the ordinate values. Overcoming the variability by collecting additional samples will not be a good approach. For Southern California, too few storms occur to create the needed sampling opportunities.

An alternative approach might be used to select a subset of representative industrial facilities from each major category. The objective is to focus limited resources on fewer samples to improve sampling methodology. In this way, the errors introduced by a poor sampling technique might be avoided, reducing the overall variability in the datasets.

This approach would be useful to estimate representative loads from each major category, but would not be useful in identifying high-risk dischargers. A large number of events could be sampled using improved sampling technology, such as flow-weighted composite samplers. The cost of such an approach could be shared and the overall cost might even be

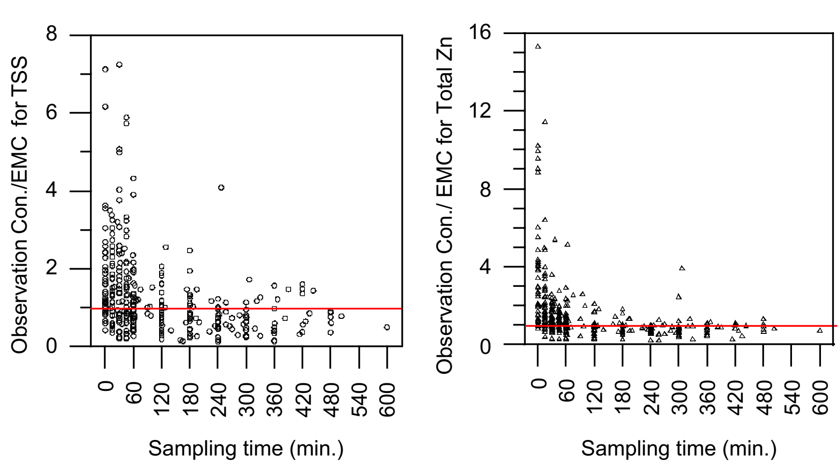

Fig. 5 – The ratio of observed concentration to EMC for TSS (left) and Total Zn (right) from UCLA First Flush Study during 2000–2001 and 2001–2002 wet seasons for highway runoff. Sampling time is the time since the beginning of measurable runoff.

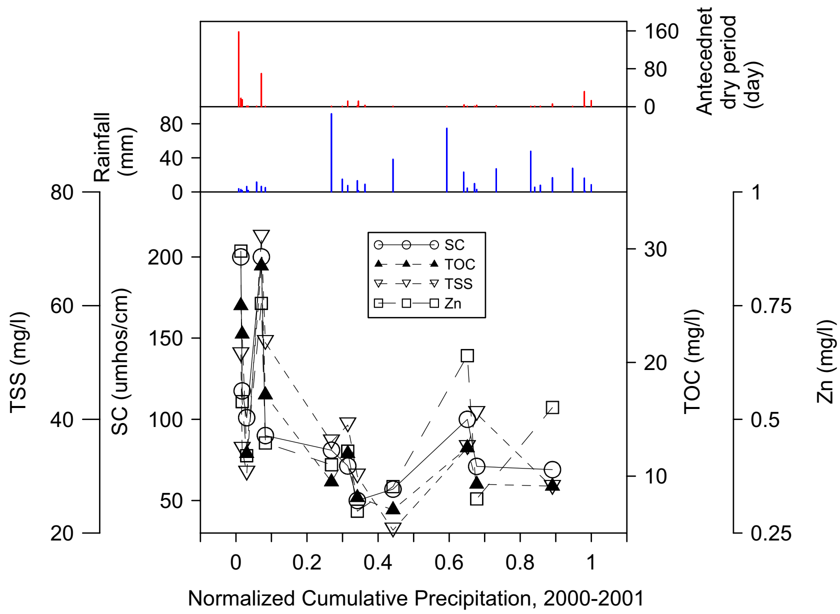

Fig. 6 – Concentrations versus normalized cumulative rainfall from Los Angeles general industrial stormwater permit monitoring during 2000–2001 wet season, illustrating the high concentrations at the end of the dry season.

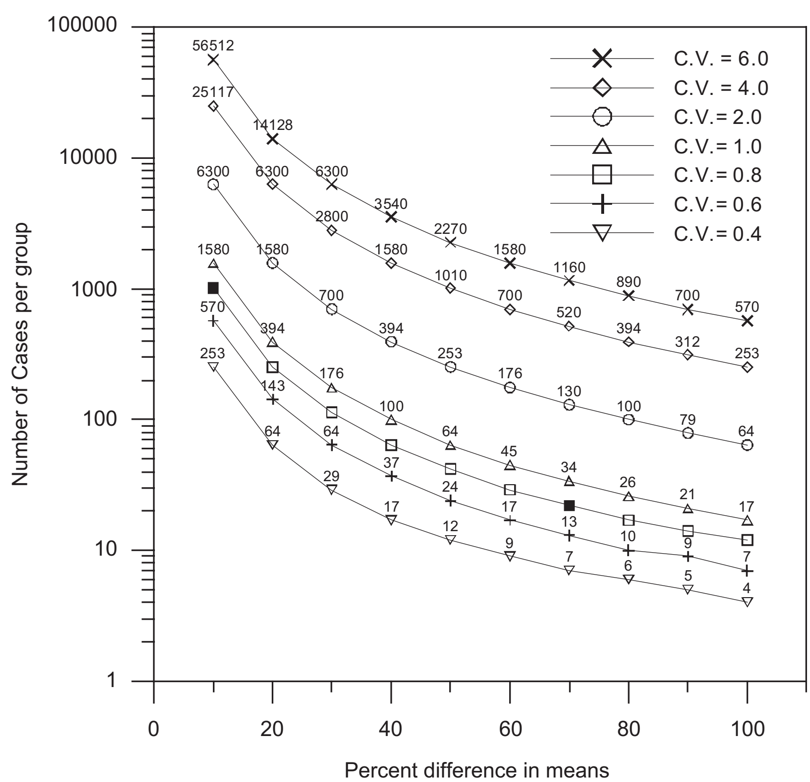

Fig. 7 – The required number of samples to detect a statistically significant difference between means as a function of relative differences in means for an α value of 0.05 and a power of 0.8.

lower than current costs. Even if only 10% of the permittees were sampled using composite samplers, the reliability of the data would be much improved and the CVs would be reduced. The remainder of the permittees could continue using grab samples or use some other program as mandated. A regulatory agency will have to develop appropriate methods for selecting sampling sites and distributing costs. Selecting a subset of each monitored category using more advanced sampling techniques is a reasonable approach and may result in lower overall cost with improved accuracy and variability.

4. Conclusions

Several stormwater monitoring programs, both national and statewide, were evaluated to determine their usefulness in identifying high dischargers and developing TMDLs. The following conclusions and recommendation are made for improving future programs:

1. Stormwater data, especially data in the Generalized Industrial Monitoring Program that were collected using grab samples, is highly variable, and typically much more variable than data observed in other monitoring programs, such as potable water or wastewater monitoring programs. Reducing the variability from experimental error and artifacts in data collection needs to be solved to improve current monitoring programs, and provide for decision making based upon collected data.
2. Statistical methods and a trained multi-layer perception neural model were used to identify relationships between water quality data and various types of industrial activity based on SIC code and landuse but were unable to detect the relationships. This was true for landuses that have known differences in water quality, and suggests that sampling methodology obscures the relationships.
3. Data from better-designed monitoring programs show that industrial landuse has the highest median concentrations of cadmium, chromium, lead, and nickel, while highways have the highest concentration of copper and zinc. In addition, concentrations of metals exceeded the US EPA stormwater benchmark level more frequently than the basic water quality parameters. In anticipation of controlling metals for TMDLs and other purposes, future permits should monitoring for metals for all industrial landuses, unless it is known that metals will not be present. Candidate metals include cadmium, chrome, copper, lead, nickel, and zinc and were included in the reviewed highway monitoring programs.
4. The timing of grab sample collection may affect results, particularly for pollutants that experience buildup and washoff, exhibiting a first flush. For these pollutants, samples collected early in the storm will have higher concentrations than the EMC and will be lower than the EMC if collected late in the storm. If grab samples are collected, they need to be collected at the appropriate time. If peak concentrations are desired, samples collected early in the storm are preferred; otherwise samples collected in the middle of the storm are more representative.
5. Most places in California including Los Angeles and Sacramento, which have distinct dry periods in the summer, had higher pollutant concentrations in early storm events than in later events, which documents the existence of a seasonal first flush. Data from Connecticut, with more uniform rainfall throughout the year, did not show a seasonal first flush. Many places in the world have distinct wet and dry periods and monitoring in these locations should consider the impacts of seasonal first flush on monitoring program goals. Samples collected early in the season will better represent maximum concentrations, but overestimate season mass loads.
6. Selecting a subset of permittees from each monitored category using more advanced sampling techniques, such as flow-weighted composite samplers, is a reasonable approach for estimating representative loads from each category and may result in lower overall cost with improved accuracy and variability.
7. We expect that our results will be helpful to planners and regulators to interpret existing datasets and programs, as well as provide recommendations for improving future programs. The results have particular value for California, but are applicable worldwide and especially in areas with Mediterranean climates.

REFERENCES

Davis, A.P., Shokouhian, M., Ni, S., 2001. Loading estimates of lead, copper, cadmium, and zinc in urban runoff from specific sources. Chemosphere 44, 997–1009.

Duke, L.D., Buffleben, M., Bauersachs, L.A., 1998. Pollutants in stormwater runoff from metal plating facilities, Los Angeles, California. Waste Manage. 18, 25–38.

US EPA, 1983. Results of the Nationwide Urban Runoff Program, vol. 1—Final Report. NTIS access number PB84-18552, December 1983, Washington, DC.

Fam, S., Stenstrom, M.K., Silverman, G., 1987. Hydrocarbons in urban runoff. J. Environ. Eng. Div. ASCE 113, 1032–1046.

Ha, H., Stenstrom, M.K., 2003. Identification of land use with water quality data in stormwater using a neural network. Water Res. 37, 4222–4230.

Han, Y., Lau, S.L., Kayhanian, M., Stenstrom, M.K., 2006. Correlation analysis among highway stormwater pollutants and characteristics. Water Sci. Technol. 53 (2), 235–243.

Khan, S., Lau, S.-L., Stenstrom, M.K., 2006. Oil and grease measurement in highway runoff—sampling time and event mean concentrations. J. Environ. Eng. ASCE 132 (3), 415–422.

Kim, L.-H., Kayhanian, M., Lau, S.-L., Stenstrom, M.K., 2005. A new modeling approach for estimating first flush metal mass loading. Water Sci. Technol. 51 (3-4), 159–167.

Lau, S.-L., Stenstrom, M.K., 2003. Catch basin inserts to reduce pollution from stormwater. Water Sci. Technol. 44 (7), 23–34.

Lee, H., Stenstrom, M.K., 2005. Utility of stormwater monitoring. Water Environ. Res. 77 (3), 219–228.

Lee, J.H., Bang, K.W., Ketchum, L.H., Choe, J.S., Yu, M.J., 2002. First flush analysis of urban storm runoff. Sci. Total Environ. 293, 163–175.

Lee, H., Lau, S.-L., Kayhanian, M., Stenstrom, M.K., 2004. Seasonal first flush phenomenon of urban stormwater discharges. Water Res. 38, 4153–4163.

Leecaster, M.K., Schiff, K., Tiefenthaler, L.L., 2002. Assessment of efficient sampling designs for urban stormwater monitoring. Water Res. 36, 1556–1564.

Olding, D.D., Stelle, T.S., Nemeth, J.C., 2004. Operation monitoring of urban stormwater management facilities and receiving subwatersheds in Richmond Hill, Ontario. Water Qual. Res. J. Can. 39 (4), 392–405.

Pitt, R., 2004. 〈http://unix.eng.ua.edu/~rpitt/Research/ms4/Final%20Table%20NSQD%20v1_1%20020704.xlsH〉.

Pitt, R., Maestre, A., Morquecho, R., 2003. Compilation and review of nationwide MS4 stormwater quality data. In: Proceedings of the 76th Water Environment Federation Technical Exposition and Conference, Los Angeles, CA.

Sansalone, J.J., Buchberger, S.G., 1997. Partitioning and first flush of metals in urban roadway stormwater. J. Environ. Eng. ASCE 123 (2), 134–143.

Stenstrom, M.K., Kayhanian, M., 2005. First flush phenomenon characterization. Report no. CTSW-RT-O5-73-02.6, California Department of Transportation, Division of Environmental Analysis, CA, USA.

Stenstrom, M.K., Silverman, G.S., Bursztynsky, T.A., 1984. Oil and grease in urban stormwaters. J. Environ. Eng. ASCE 110 (1), 58–72.

Turner, B.G., Boner, M.C., 2004. Watershed monitoring and modeling and USA regulatory compliance. Water Sci. Technol. 50 (11), 7–12.