

Evaluation of street sweeping effectiveness as a stormwater management practice using statistical power analysis

Joo-Hyon Kang and Michael K. Stenstrom

ABSTRACT

Although street sweeping is commonly regarded as a cost-effective stormwater best management practice, there is little quantitative evidence that street sweeping directly improves runoff water quality. In this paper, several previous street sweeping studies were re-evaluated using statistical power analysis. Two-group, independent-sample one-sided t-test power analyses were performed using log-transformed event mean concentrations (EMCs) of total suspended solids, suspended sediment concentration or chemical oxygen demand. The effect size between the two groups was estimated using the sweepers' pick-up efficiency, which showed that the failure to detect the difference between mean EMCs of the two sample groups (i.e., unswept and swept groups) is likely due to limited sample numbers. Too few samples, which also resulted in a high coefficient of variation, were analysed to detect the likely difference between swept and unswept observations. In addition, the temporal gap between street sweeping and subsequent storm events was not controlled to improve statistical power.

Key words | best management practices (BMPs), statistical power analysis, stormwater, street sweeping, urban runoff

Joo-Hyon Kang

Post-Doctoral Researcher,
Saint Anthony Falls Laboratory,
University of Minnesota at Twin Cities,
#2 Third Avenue S.E.,
Minneapolis,
MN 55414,
USA
E-mail: joohyon.kang@engineering.ucla.edu

Michael K. Stenstrom

Department of Civil and Environmental
Engineering,
University of California at Los Angeles,
5714 Boelter Hall,
Los Angeles,
CA 90095-1593,
USA
E-mail: stenstro@seas.ucla.edu

INTRODUCTION

Street sweeping is typically practised to pick up large materials such as litter (defined as anthropogenic materials such as plastic bags and cigarette butts etc.) and natural origin debris (gravel and vegetation) to improve aesthetics. Currently, street sweeping is being considered an important BMP for stormwater management; however, early studies (Sartor & Boyd 1972; US EPA 1983; Sartor & Gaboury 1984; Smith 2002) claimed that street sweeping was ineffective at reducing pollutant concentrations in stormwater runoff due to low efficiency of fine particle removal. These early reports are fuelling a debate over the value of street sweeping or the need for additional end-of-pipe treatment devices (Walker & Wong 1999; Schilling 2005).

Sweeping efficiency varies depending on sweeping frequency, sweeper operating speed, sweeping technologies,

operator care and initially deposited sediment load (Sutherland & Jelen 1997; US EPA 1999; Curtis 2002). The ability of sweepers to pick up fine particles is the most important factor to prevent pollutant discharge in subsequent storm events. The early street sweeping studies performed as a part of the Nationwide Urban Runoff Program (NURP) suggested that street sweeping pick-up efficiency for particles below 100 μm was less than 20% (US EPA 1983), where many if not most of the pollutants such as heavy metals are found. However, street sweeping is still considered effective by others (Sutherland & Jelen 1997; MWH Americas 2002) to improve stormwater quality when the right sweeper technologies and operation methods are used. Although there have been numerous studies to evaluate sweeping efficiencies, little evidence has been documented that street sweeping directly improves

stormwater quality, despite the intuitive belief that the large mass of swept material would have become water pollution.

In this paper, previous street sweeping studies were reviewed and critically assessed. Postulating that street sweeping causes reduction in pollutant event mean concentrations (EMCs) of end-of-pipe runoff, statistical power analysis was conducted using 15 datasets for total suspended solids (TSS), suspended sediment concentration (SSC) or chemical oxygen demand (COD) in 13 different locations from four previous street sweeping studies. Two-group, independent-sample, one-sided *t*-tests were used for the power analysis. The sample groups of EMCs under unswept and swept conditions were considered the control group and experimental group, respectively. The statistical power values of individual sites were calculated to assess the probability of rejecting the hypothesis that street sweeping does not cause reduction in EMCs. The goal of this study was to assess the statistical confidence of the previous street sweeping studies and to discuss important considerations in evaluating the effectiveness of street sweeping as a stormwater BMP.

METHODS

In order to evaluate the performance of street sweeping as a stormwater BMP, datasets of end-of-pipe EMCs under swept and unswept conditions (unswept and swept sample groups) can be compared. The paired basin approach (unswept and swept basins located near each other are simultaneously observed in a test period) or serial basin approach (unswept and swept modes in a test basin are sequentially observed) can be used for the comparison study (i.e. two-group research) (US EPA 1983). The difference between the two sample groups might be difficult to detect due to low removal rate of water quality pollutants, or the limited number of observations may obscure statistical conclusions because the variation within each group is relatively high when compared to the difference between sample groups. The high variability occurs because of regional and seasonal differences, random nature of meteorological conditions (i.e., rainfall intensity, rainfall duration and dry period between storms) and measurement errors.

Probability distribution of EMCs for power analysis

Statistical power analysis was performed using two-group, independent-sample one-sided *t*-tests. The goal was to evaluate statistical reliability of the earlier street sweeping studies including the NURP study, which is a canonical reference for criticising the effectiveness of street sweeping. Control and experimental groups were the EMCs observed under unswept and swept conditions (unswept and swept sample groups). Because EMCs within a specific location typically follow log-normal distribution (US EPA 1983), log-transformed EMCs ($\ln[\text{EMC}]$ s) were used as the variable in the power analysis to meet the requirement of population's normality for the *t*-test. The probability distribution of a control group can be characterised by mean $\mu_{\ln[x]}$ and standard deviation $\sigma_{\ln[x]}$ on the basis of $\ln[\text{EMC}]$, which were calculated by substituting the values of median (Md) and coefficient of variation (CV) of EMCs into the relationships of log-normal distribution as follows (Aitchison & Brown 1957):

$$\mu_{\ln[x]} = \ln[\text{Md}] \quad (1)$$

$$\sigma_{\ln[x]} = \sqrt{\ln[1 + \text{CV}^2]} \quad (2)$$

where $\mu_{\ln[x]}$ is the mean of log-transformed data, n is the number of data, Md is the median of data, $\sigma_{\ln[x]}$ is standard deviation of log-transformed data, and CV is coefficient of variation of data. It was also assumed that the log-transformed EMCs for swept conditions (experimental group) had smaller $\mu_{\ln[x]}$ but the same $\sigma_{\ln[x]}$, as opposed to a control (unswept) group.

Effect size estimation

The effect size (ES) between two groups is defined as the difference between means of two groups divided by pooled standard deviation. The ES reflects the strength of experimental effects in practice. In this study, the difference between means of a control (unswept) and experimental (swept) group was approximated using the sweeper's pick-up efficiency. That is, assuming that street sweeping is always performed right before a storm and ignoring wet deposition (i.e., no build up occurs after street sweeping until the end of a subsequent storm), the EMC can be

proportionally related to the initial pollutant mass available on the road surface. The pooled standard deviation was approximated by $\sigma_{\ln[x]}$, assuming the control and experimental group have the same standard deviation. Because a sweeper directly reduces the pollutant mass available on the road surface, the ES can be formulated as follows:

$$\begin{aligned} ES &= \frac{\mu_{\text{unswept}} - \mu_{\text{swept}}}{\sigma_{\ln[x]}} = \frac{\ln[Md_{\text{unswept}}] - \ln[Md_{\text{swept}}]}{\sigma_{\ln[x]}} \\ &= \ln(Md_{\text{unswept}}/Md_{\text{swept}})^{1/\sigma_{\ln[x]}} \\ &= \ln[Md_{\text{unswept}}/(1 - R_s)Md_{\text{unswept}}]^{1/\sigma_{\ln[x]}} \\ &= \ln[1 - R_s]^{-1/\sigma_{\ln[x]}} = \ln[1 - R_s]^{-1/\sqrt{\ln[1 + CV^2]}} \end{aligned} \quad (3)$$

where μ_{unswept} and μ_{swept} are means of $\ln[EMC]$ s under unswept and swept conditions, respectively, Md_{unswept} and Md_{swept} are medians of EMCs under unswept and swept conditions, respectively, and R_s is the reduction rate of road deposited pollutant by street sweeping. The value of R_s can be estimated using sweeper's pick-up efficiency and the percentage of pollutants deposited on the sweeper's path (typically on the curb area) as opposed to total catchment area as follows:

$$R_s = f_{\text{curb}} \cdot E_p \quad (4)$$

where f_{curb} is mass fraction of sediments deposited in curb areas, E_p is sweeper's pick-up efficiency. Table 1 shows the pick-up efficiencies of mechanical broom sweepers and

vacuum assisted sweepers for different particle size ranges (Sartor & Boyd 1972; Breault *et al.* 2005). Because the runoff particles are typically smaller than 500 μm and rarely larger than 2000 μm (Walker & Wong 1999; Li *et al.* 2006), the value of E_p can be calculated using the weight-averaged pick-up efficiency of five different particle size ranges less than 2,000 μm (i.e., <43 μm , 43–104 μm , 104–246 μm , 246–840 μm , 840–2,000 μm) as follows:

$$R_s = f_{\text{curb}} \cdot \frac{\sum_{i=1}^m w_i e_i}{\sum_{i=1}^m w_i} \quad (5)$$

where m is the number of particle size ranges less than 2,000 μm , w_i is weight fraction of particles in i th size range, and e_i is sweeper's pick-up efficiency for particles in i th size range. Applying 0.8 of f_{curb} (Sartor & Boyd 1972), R_s for solids (TSS or SSC) for mechanical broom sweepers and vacuum assisted sweepers were calculated as 38% and 59%, respectively. Similarly, E_p for COD can be estimated using the fraction of COD in different particle size ranges (Table 1) and calculated as 24% and 59% for mechanical broom sweepers and vacuum assisted sweepers, respectively. Mean, median, and standard deviation of EMCs reported in each study were used to calculate the ES.

Calculations of statistical power and sample size for the previous studies

Our power analyses included 15 studies (i.e. 15 EMC datasets) from 13 different locations: 10 datasets for TSS from different NURP sites (US EPA 1983), datasets for COD and TSS in a single site from Irish *et al.* (1995), one SSC dataset from Smith (2002), and datasets for TSS and SSC in a single site from Martinelli *et al.* (2002). Using 0.05 of significance level (α) and statistical parameters obtained from the previous studies, values of statistical power (β) and required sample number at $\beta = 0.8$ were calculated for each site.

RESULTS AND DISCUSSION

Evaluation of the previous studies

Figure 1 shows the functional relationships between β and sample number per group for different values of ES at

Table 1 | Pickup efficiency of mechanical broom sweepers

Size range of particles (μm)	Fraction (%)	Pollutant fraction (mg/kg TS)			Pickup efficiency for particles (%)	
		COD	Cu	Zn	M	V*
0–43	5.9	261,077			15	81
43–104	9.7	314,801	201	791	20	60–78
104–246	27.8	30,267	98	444	48	57–78
246–840	24.6	35,859	95	302	60	78
840–2,000	7.6	40,178	376	1,582	66	78
>2,000	24.4	6,674	131	93	79	90
Weighted Average (<2,000 μm)					47	74

*By Breault *et al.* (2005), mean of two experiments.

Adapted from Sartor & Boyd 1972, except vacuum sweeper efficiency. Mean values of residential, commercial and industrial streets. M-mechanical broom sweeper, V-vacuum assisted sweeper.

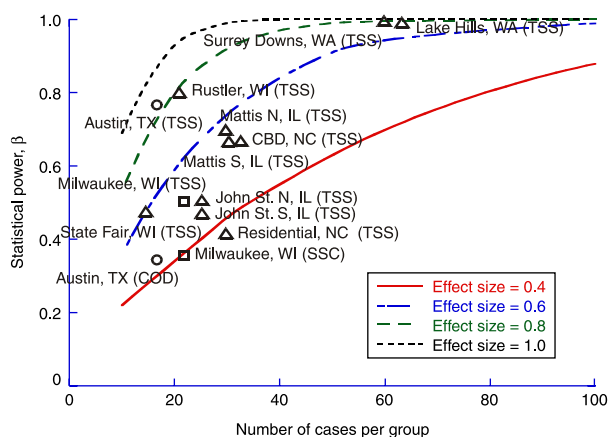


Figure 1 | Functional relationships between β and sample number per group for different values of ES at $\alpha = 0.05$ (one-sided t -test).

$\alpha = 0.05$, which was created using two-group independent-sample one-sided t -tests. Data points corresponding to the 15 studies (15 EMC data sets) were also superposed on the graphs. For the NURP studies, in general 70 samples per group are required for 0.8 or larger β . Using the same statistical procedures, the functional relationship between number of samples per group and the ES can be obtained for $\beta = 0.8$ and $\alpha = 0.05$, and is shown in Figure 2. The positions on the graph corresponding to the 15 studies are also designated in the figure.

Figure 1 shows that only four of the 15 previous studies had the statistical power large enough to detect a difference in water quality from street sweeping. The four studies

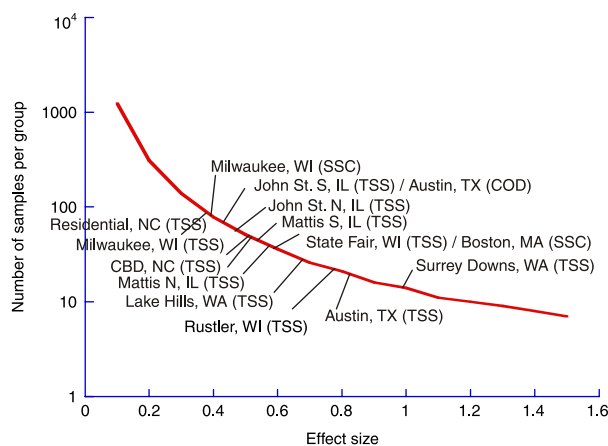


Figure 2 | Functional relationship between number of samples per group and ES (one-sided t -test with $\alpha = 0.05$ and $\beta = 0.8$).

(Austin, TX (TSS) [$\beta = 0.77$]; Rustler, WI (TSS) [$\beta = 0.80$]; Surrey Down, WA (TSS) [$\beta = 0.98$]; Lake Hills, WA (TSS) [$\beta > 0.99$]) were likely to find a statistically significant difference, if one existed. Of the four studies with sufficient statistical power, however, only one study (Austin, TX (TSS), Irish *et al.* 1995) found a difference in water quality, and they reported a reduction in TSS with high confidence level using t -statistics ($|t| = 3.53 > t_{0.01,\infty} = 2.33$). They did not find a difference in other water quality parameters (i.e. COD, metals, nutrients), but this is not surprising because the statistical power for these other water quality parameters is lower (e.g. $\beta = 0.34$ for COD), reducing the probability of detecting the difference, even if it has existed. Despite high statistical power, the other three studies (all from NURP studies) observed no statistical differences between EMCs under unswept and swept conditions. Several researchers noted that the frequency of sweeping affects the street sweeping efficiency (Sutherland & Jelen 1997; US EPA 1999; Curtis 2002) and this might be the cause of the negative findings. Note that the values of ES calculated in Table 3 are the maximum values possible assuming no pollutant buildup between sweeping and the following storm event. This assumption assures that the maximum value of ES is assumed, but the ES could be much smaller if there is a large temporal gap between sweeping performance and the subsequent storm event (the time between sweeping and the following storm event is not reported in the NURP documents). This possibility must be considered when evaluating the reasons for no improvement in water quality for the three examples with high ES. The effect of frequency on sweeping efficiency is discussed more in the next section.

Figure 2, or similar figures generated in the same way but with different values of α and β , can be used to determine the number of samples required to detect a significant difference. Power analysis for other pollutants such as COD or heavy metals associated with the particulates can be performed in a similar way using the information of pollutant fractions. Because of the lower removal rates of smaller particles by the sweepers, the values of ES for COD or other pollutants are usually smaller than the values of ES for TSS or SSC, and will require a larger number of observations to detect the difference between the EMCs under swept and unswept conditions.

Ideally, an exercise to determine the required number of samples should be performed before undertaking an experimental program. This can be easily done if the variability of the measurements (i.e., for this case, the CV of the pollutant EMCs in the runoff) is known.

Among the 15 previous study results contained in Figure 2, the four studies with sufficient statistical power discussed earlier correspond to the largest four ES values and the smallest four required sample numbers (<30). As stated earlier, one of the four studies (Austin, TX (TSS)) matched our power analysis well but the other three studies (Rustler, WI (TSS), Surrey Down, WA (TSS), Lake Hills, WA (TSS)) reported negative results. This can be explained by overestimated ES used in the power analysis. Figure 2 implies the significant impact of the overestimated ES on the statistical power; that is, linear decrease in ES results in exponential increase in the required number of samples per group to keep the same level of statistical power.

Our observation of the previous studies, with one exception (Irish *et al.* 1995), is that their statistical designs were not sufficiently robust to detect improvements in water quality due to sweeping.

Therefore the previous references to the NURP studies on street sweeping ineffectiveness on water quality should be reevaluated. Future references should be made as “unable to detect differences due to statistical design or limited numbers of samples.”

Effect of street sweeping

The inability of previous studies to detect differences in water quality from swept and unswept conditions is in part due to the limited number of observations, but also to other phenomena, such as sweeping frequency, buildup rate and rainfall characteristics. This section discusses how these parameters affect the statistical power and how a study design can be improved.

The ES is strongly affected by the differences in the two means, and for street sweeping this is determined by the sweeper efficiency as well as the sweeping schedule. More efficient sweepers should produce a greater difference in means which should be easier to detect. The sweeping schedule is also a critical factor that affects the ES. Pollutant

mass on street surfaces builds up by different mechanisms, and little improvement in runoff water quality will be observed if buildup occurs between street sweeping and the sampled storm event. Therefore, the ES should be chosen as a function of the timing and frequency of street sweeping. More conservative ES (i.e., smaller ES) is required for infrequent sweeping or in high buildup areas.

Figure 3 shows two conceptual examples of pollutant buildup on streets, one with and one without sweeping. The mass of pollutants eventually reaches a steady state and various models have been used to describe the buildup (Southerland & Jelen 1993; Charbeneau & Barrett 1998). As the pollutant mass on street surface reaches steady state, the mass that would have accumulated is released to other environmental sinks, and can become an airborne pollutant, increasing ambient particle concentrations (e.g., PM_{10} , $PM_{2.5}$), or can be transported to adjacent or distant locations by advection and dispersion

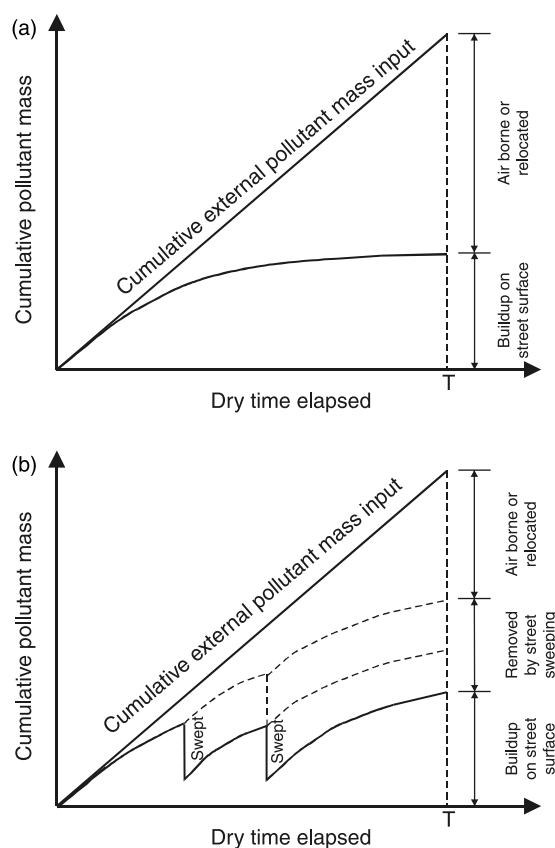


Figure 3 | Destination of external pollutant sources in street surface system: (a) under unswept condition; (b) under swept condition.

mechanisms. If rainfall occurs at time T , the swept condition (Figure 3(b)) has the same amount of deposition as the unswept condition (Figure 3(a)) which will result in the same runoff water quality (i.e., no observed improvement from sweeping), but pollutant mass released to other environmental sinks is reduced.

Street sweeping may provide benefits in addition to water quality improvements. This example shows that an evaluation of street sweeping benefits, solely on the basis of water quality and litter mass recovered, does not capture all the benefits. Long-term studies to evaluate the changes in PM_{10} and $PM_{2.5}$ for routinely swept and unswept streets, to the best of our knowledge, have not been performed, but should be, to fully assess the benefits of sweeping.

Design of experimental study

To design a monitoring program for sweeping efficiency, the number of samples to be collected should be determined using the previous knowledge of the EMCs of the pollutants to be monitored, such as mean and variance, as well as a specified statistical power and significant level. The number of samples to be collected should be determined as a function of the street sweeping technique, expected pollutant build-up, and street sweeping protocols such as frequency and coverage. More efficient sweeping technologies such as vacuum assisted sweepers have higher pickup efficiency, resulting in larger effect size for power analysis, and will require fewer samples.

When a street sweeping study is focused on air pollution issues such as abating ambient particle concentrations, long-term investigations are recommended because a single sweeping can not provide dramatic reduction in airborne pollutant mass. For example, several previous studies showed that little instantaneous reduction or even a short-lived increase in ambient particle concentrations will be observed after street sweeping (Fitz 1998; Kuhns *et al.* 2003; Chang *et al.* 2005). This can be inferred from Figure 3(b), where the distance between the dotted line and the 45° diagonal is the mass of airborne particles. The time required between sweeping and the time to achieve equilibrium of pollutant mass on the surface can be significant. A simple measurement of airborne particle mass

just before and just after sweeping does not measure a longer term reduction in particle mass that can occur after equilibrium is achieved.

CONCLUSIONS

A statistical power analysis was performed to re-evaluate the effectiveness of street sweeping using 15 EMC datasets from the four street sweeping studies previously performed by other researchers. End-of-pipe EMCs for TSS, SSC or COD from 13 different locations were used in logarithmic form for two-group independent-sample one-sided t -test power analysis. The probable difference between EMCs under unswept and swept conditions (two sample groups) was estimated using the pickup efficiency of the street sweeper and its coverage area.

The results show that, in general, the NURP study and other previous street sweeping studies used too few samples, resulting in low statistical power and inability to detect the benefits of street sweeping to water quality. Only four of the 15 EMC datasets re-evaluated had sufficient numbers of samples to produce high power values ($\beta > \sim 0.8$). Three of the four datasets with high statistical power failed to detect a water quality improvement after sweeping for TSS. This may have resulted from an inability to detect large particles in the runoff, which has been suggested by others (Gray *et al.* 2000), or the build up mechanism, which reduces effect size. The other dataset with high statistical power (Irish *et al.* 1995) detected TSS reduction after street sweeping. Their results matched our power analysis well; that is, after street sweeping, measurable reduction was observed for TSS with high statistical power and no reduction was observed for COD with low statistical power.

The contribution of street sweeping to environmental quality in urban areas should not be underestimated because of previous studies, which had insufficient statistical power to detect water quality improvements, had they existed. New studies using modern sweeping technology and better statistical designs to detect probable differences should be performed. The statistical design can be optimized to create the smallest number of samples to detect the expected difference and avoiding high analytical costs.

REFERENCES

- Aitchison, J. & Brown, J. A. C. 1957 *The Lognormal Distribution*. Cambridge University Press.
- Breault, R. F., Smith, K. P. & Sorenson, J. R. 2005 *Residential street-dirt accumulation rates and chemical composition, and removal efficiencies by mechanical- and vacuum-type sweepers, New Bedford, Massachusetts, 2003–04*. Scientific Investigations Report 2005-5184, U.S. Department of the Interior, U.S. Geological Survey, Reston, Virginia.
- Charbeneau, R. J. & Barrett, M. E. 1998 *Evaluation of methods for estimating stormwater pollutant loads*. *Water Environ. Res.* **70**(7), 1295–1302.
- Chang, Y.-M., Chou, C.-M., Su, K.-T. & Tseng, C.-H. 2005 *Effectiveness of street sweeping and washing for controlling ambient TSP*. *Atmos. Environ.* **39**, 1891–1902.
- Curtis, M. C. 2002 *Street sweeping for pollutant removal*. Watershed Management Division, Department of Environmental Protection, Montgomery County, Maryland.
- Fitz, D. R. 1998 *Evaluation of street sweeping as a PM₁₀ control method*. Final Report, Mobile Source Air Pollution Review Committee, South Coast Air Quality Management District.
- Gray, J. R., Glysson, G. D., & Conge, L. M. 2000 *Comparability of total suspended solids and suspended-sediment concentration data*. Water-Resources Investigations Report 00-4191, U.S. Department of the Interior, U.S. Geological Survey, Reston, Virginia.
- Irish, L. B. Jr., Lesso, W. G., Barrett, M. E., Malina, J. F. Jr. & Charbeneau, R. J. 1995 *An evaluation of the factors affecting the quality of highway runoff in the Austin, Texas area*. Technical Report, CRWR 264, Center for Research in Water Resources, Bureau of Engineering Research, University of Texas, Austin, TX.
- Kuhns, H., Etyemezian, V., Green, M., Hendrickson, K., McGown, M., Barton, K. & Pitchford, M. 2003 *Vehicle-based road dust emission measurement-Part II: effect of precipitation, wintertime road sanding, and street sweepers on inferred PM₁₀ emission potentials from paved and unpaved roads*. *Atmospheric Environ.* **37**, 4573–4582.
- Li, Y., Lau, S.-L., Kayhanian, M. & Stenstrom, M. K. 2006 *Dynamic characteristics of particle size distribution in highway runoff: implications for settling tank design*. *J. Environ. Eng.* **132**(8), 852–861.
- Martinelli, T., Waschbusch, R., Bannerman, R. & Wisner, A. 2002 *Freeway Sweeping to Reduce Runoff Pollutants*. Wisconsin Department of Transportation R&D Program, Madison, Wisconsin.
- MWH Americas, Inc. 2002 *Anchorage OGS street sweeping as storm water controls Performance analysis*. WMP Apr02002, Watershed Management Program, Project Management and Engineering, Department of Public Works, Municipality of Anchorage, Alaska.
- Sartor, J. D. & Boyd, G. B. 1972 *Water pollution, aspects of street surface contaminants*. EPA-R2-72-081, US EPA.
- Sartor, J. D. & Gaboury, D. R. 1984 *Street sweeping as a water pollution control measure: lessons learned over the past ten years*. *Sci. Total Environ.* **33**, 171–183.
- Schilling, J. G. 2005 *Street sweeping—Report No. 1, State of the practice*. Prepared for Ramsey-Washington Metro Watershed District. <http://www.rwmwd.org>. North St. Paul, Minnesota.
- Smith, K. P. 2002 *Effectiveness of three best management practices for highway-runoff quality along the Southwest Expressway, Boston, Massachusetts*. Water-Resources Investigations Report 02-4059, U.S. Department of the Interior, U.S. Geological Survey, Northborough, Massachusetts.
- Sutherland, R. C. & Jelen, S. L. 1993 *Simplified Particulate Transport Model-Users Manual*. Kurahashi and Assoc., Inc., Tigard, Oregon.
- Sutherland, R. C. & Jelen, S. L. 1997 *Contrary to conventional wisdom, street sweeping can be an effective BMP*. In: James, W. (ed.) *Advances in Modeling the Management of Stormwater Impacts (Vol. 5)*. Computational Hydraulics International, Cuelph, Ontario, pp. 179–190.
- United States Environmental Protection Agency 1983 *Results of the nationwide urban runoff program*. Vol I- Final Report, Water Planning Division, WH-554, Washington, D.C.
- United States Environmental Protection Agency 1999 *Storm Water Management Fact Sheet—Dust Control*. EPA 832-F-99-003, Office of Water, Washington, D.C.
- Walker, T. A. & Wong, T. H. F. 1999 *Effectiveness of street sweeping for stormwater pollution control*. Technical Report, Report 99/8, Cooperative Research Centre for Catchment Hydrology, Australia.