Sampling Issues in Urban Runoff Monitoring Programs: Composite versus Grab

Jiun-Shiu Ma[1]; Joo-Hyon Kang[2]; Masoud Kayhanian[3]; and Michael K. Stenstrom[4]

Abstract: Storm-water monitoring generally uses flow-weighted automatic composite samplers to collect a representative sample of an entire storm event. Automatic samplers are convenient but unfortunately they can be expensive, especially for temporary sampling needs or for short-term research projects. An alternative method is to use a series of grab samples. This paper examines the accuracy of event mean concentrations (EMCs) and mass first flush ratios calculated from a finite number of grab samples, and compares them to results from flow-weighted automatic samples. Both sampling techniques were evaluated using data collected from a three-year investigation of three highway sites. A large number of grab samples is needed to approach the accuracy and precision of flow weighted composite samples, and 30 grab samples per storm event generally estimated the EMCs within 20% average error. To detect a first flush, it is necessary to take even more grab samples or to adjust the timing of the sample collection toward the beginning of the storm. The superiority of automatic sampling for estimating EMCs for constituents compatible with automatic sampling is demonstrated.

DOI: 10.1061/(ASCE)0733-9372(2009)135:3(118)

CE Database subject headings: Stormwater management; Composite materials; Regression analysis; Sampling; Runoff; Urban areas.

Introduction

The quality of storm-water runoff is typically represented by the event mean concentrations (EMCs) and mass loading diagrams or mass first flush (MFF) ratios to quantify the first flush. Both can be important parameters to characterize nonpoint sources and establish best management practice (BMP) strategies for decision makers and engineers. EMCs have been extensively used to characterize storm-water pollutant loads (USEPA 1983; Charbeneau and Barrett 1998; Carleton et al. 2001). The EMC, by its name, represents an average concentration of a specific pollutant contained in runoff throughout a storm event. The EMCs differ from grab sample concentrations in that they estimate the entire storm event, as opposed to a single point of time in the storm event. The EMCs can be multiplied by runoff volumes to calculate total discharge pollutant masses. The MFF ratio (Ma et al. 2002) is a dimensionless parameter to quantify the degree of mass first flush, which occurs when a greater fraction of the pollutant mass is discharged during the initial part of a storm. As opposed to the EMCs, the MFF ratios characterize time varying discharge behavior of a specific pollutant in a storm and can be used to estimate the magnitude of the initial storm pollutant mass discharge.

An ideal method to obtain the EMC of a storm event is to collect the whole volume of runoff and measure pollutant concentrations, which is physically unrealistic. Alternatively, the continuous measurement of both pollutant concentrations and flow rates in a storm event can provide accurate EMC estimates, but may also be impractical because the concentrations must be measured with automated analyzers, which are either expensive, difficult to maintain, or not available. For practical reasons, the EMCs are usually determined using flow-weighted automatic samplers, which pace sample collection with runoff rate to obtain the EMC (USEPA 2002). If unavailable or if time varying data are required, such as the data needed to characterize the first flush, discrete water samples along with continuous flow data must be used (Charbeneau and Barrett 1998; Khan et al. 2006). Each grab sample represents an instantaneous concentration within a storm event, and the EMC is calculated using these instantaneous concentrations along with the representative flow volume (USEPA 2002; Gulliver and Anderson 2008). A reasonable EMC estimator is the flow-weighted average of these instantaneous concentrations (Charbeneau and Barrett 1998; Wu et al. 1998). Flow-weighted discrete samples can also be used to approximate the continuous profile of pollutant mass discharge, which is used to calculate MFF ratios. More frequent sampling in the earlier part of a storm is usually needed to better estimate the first flush, due to the rapid change in runoff quality in the initial runoff (Stenstrom and Kayhanian 2005).

Automatic samplers are often preferred because they can be operated unattended and can be programmed to adjust to dynamically changing runoff including the time to initiate and end sampling as well as pacing the sample volume collection with runoff flow. The samplers collect a large number of discrete samples, which are then combined into one composite sample. The samplers are programmed to collect a small, constant volume of runoff at every time interval or volume of runoff. In this way, the

[1]Water Resource Manager, Fuscoe Engineering, 16795 Von Karman, Ste. 100, Irvine, CA 92606.

[2]Research Professor, Dept. of Environmental Science and Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju 500-712, Korea (corresponding author). E-mail: joohyon@umn.edu

[3]Associate Director, Center for Environmental and Water Resource Engineering, Dept. of Civil and Environmental Engineering, Univ. of California, Davis, CA 95616.

[4]Professor, Department of Civil and Environmental Engineering, Univ. of California at Los Angeles, Los Angeles, CA 90095-1593.

Note. Discussion open until August 1, 2009. Separate discussions must be submitted for individual papers. The manuscript for this paper was submitted for review and possible publication on July 25, 2007; approved on October 28, 2008. This paper is part of the ***Journal of Environmental Engineering***, Vol. 135, No. 3, March 1, 2009. ©ASCE, ISSN 0733-9372/2009/3-118–127/$25.00.

samplers can produce unequal time series discrete samples or one single composite sample that represents the EMC.

Automatic samplers are limited for many pollutants because of carry over between samples or the storage time that occurs in the collection vessel. Automatic samplers are not recommended for pH and dissolved oxygen concentration, fecal indicator bacterial (Gulliver and Anderson 2008), and particle size distribution (Li et al. 2005) because of the changes that can occur during sample storage. They are generally not used for oil and grease or hydrocarbons due to carry over during successive sampling or adsorption on tubing and bottle surfaces. They may also introduce artifactual toxicity due to contact with tubing or the collection vessel. The general applicability of automatic samplers for specific pollutants is beyond the scope of this paper, but needs to be considered in designing a monitoring program.

Even though an automatic sampler is collecting a composite sample, the EMC can still mathematically be viewed as a result from a series of discrete samples. The number of discrete samples depends on the sample volume being collected and the discharge volume criteria that determine sampling times. It is intuitively correct that a larger number of discrete samples will produce a more reliable estimate of the EMCs and MFF ratios. It has also been speculated that significant errors can be introduced into EMC calculations by insufficient sample size, especially when the concentrations of discrete samples can vary two orders of magnitude during a storm event, which was routinely observed during the field study that supported this work. The error and sensitivity of EMC estimates to sample size has not been previously explored.

The goal of this paper is to compare the reliability of the EMCs and MFF ratios collected with automatic samplers and those calculated from a varying number of grab samples. Chemical oxygen demand (COD), which is one of the most commonly used water quality parameters, is used as a representative parameter. COD was selected as a surrogate water quality parameter because COD is the best indicator of oxygen demand, relatively inexpensive to analyze, has low measurement error, and is not subject to interferences. In addition, our previous monitoring study revealed strong correlations with other highway stormwater quality parameters; therefore, other parameters can be predicted from COD (Han et al. 2006).

The EMC and MFF ratios are estimated from different sampling strategies: randomly timed, equally timed, equal-rainfall depth, equal-discharge volume, and first flush enhanced samplings. One-minute interval flow data for 35 storm events monitored from three highway runoff monitoring sites during the 1999–2001 wet seasons were used to simulate the five sampling strategies. To obtain reference values of the COD of discrete samples, a regression model (Ma 2002) previously developed for the monitoring sites was used. The COD of grab samples was simulated by adding random sampling errors to the COD predicted by the regression model, and the EMCs and MFF ratios were calculated by integrating the flow-weighted pollutant discharge rate. The simulated EMCs and MFF ratios are compared to the reference values of EMC and MFF ratios, respectively, to assess the accuracy of each sampling strategy.

Methodology

Data Sources

A large suite of grab samples and composite samples were collected for 35 storm events in the three highway runoff monitoring sites (Sites 1, 2, and 3) located in west Los Angeles from 2 years of observation (1999–2001). Generally, 12 samples were manually collected per storm event at 15 min intervals in the first hour of runoff and at 1 h intervals for the rest of the storm. Flow-weighted composite samples were collected using automated samplers (American Sigma 900Max, Ontario, Canada). The hyetograph and hydrograph of each storm were also obtained using tipping bucket rain gauges and flow meters associated with the composite samplers. More detailed site descriptions and monitoring methodologies are available elsewhere (Stenstrom and Kayhanian 2005).

Estimation of EMCs and MFF Ratios Using Discrete Samples

Mathematically, the EMCs can be defined as total pollutant mass (M) discharged during an event divided by total volume (V) discharge of the storm event as follows:

$$\mathrm{EMC} = \frac{M}{V} = \frac{\int C(t)Q(t)dt}{\int Q(t)dt} \tag{1}$$

where $C(t)$=continuous function of time that represents the pollutant concentration curve; and $Q(t)$ is also a continuous function of time that represents the storm-water discharge flow rate curve. However, in practice, $Q(t)$ and $C(t)$ are not continuous functions but discrete measurements of $Q(t)$ and $C(t)$ at various points in time. It is necessary to estimate the EMC from the discrete values, and if the concentration and the discharge rate are measured at the same time, the EMC can be estimated as

$$\mathrm{EMC} = \frac{\Sigma_i c_i q_i \Delta t_i}{\Sigma_i q_i \Delta t_i} \tag{2}$$

where q_i and c_i=measurements for the discharge rate (L^3/T) and pollutant concentration (M/L^3) in the ith interval and Δt_i =length of time in the ith interval (T). From the point of view of approximating the continuous functions in Eq. (1), more accurate approximations can be obtained with Eq. (2) when using more measurements. When the discharge volumes are viewed as weights, Eq. (2) becomes the discharge volume-weighted average throughout the storm event

$$\mathrm{EMC} = \sum_i w_i c_i \tag{3}$$

$$w_i = \frac{q_i \Delta t_i}{\Sigma_i q_i \Delta t_i} \tag{4}$$

where w_i=flow weight; and $\Sigma_i w_i = 1$. In practice, a common situation is for a few concentration measurements to be made while a large number of discharge measurements are made. Generally, there are many fewer concentration measurements because concentration measurements are more expensive and time consuming; discharge measurements can be easily and automatically obtained by the composite sampler's flow measuring sensors. For most situations, we have to adjust the weights for each concentration measurement in Eq. (3). One reasonable way to adjust the weights is to use the discharge volume. One approach (Charbeneau and Barrett 1998) splits the discharge volume from the midpoint between two consecutive concentration measurements. Fig. 1 conceptualizes this approach. If j is the number of the flow measurements in the ith interval of the concentration measurement, the adjusted weight can be written as

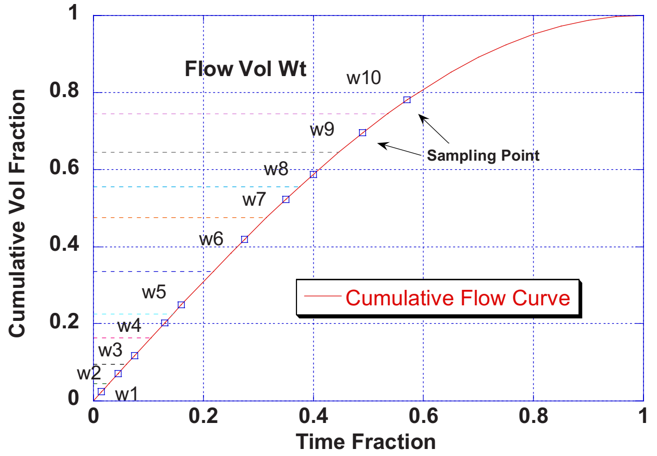

Fig. 1. Determination of flow weights (w1 to w10) for grab samples

$$w_i = \frac{V_i}{\Sigma_i V_i} = \frac{\Sigma_j q_{ij} \Delta t_{ij}}{\Sigma_i \Sigma_j q_{ij} \Delta t_{ij}} \tag{5}$$

where V_i=discharge volume corresponding to the ith concentration measurement (L^3), q_{ij}=jth flow measurement in the ith interval of the concentration measurement (L^3/T); and Δt_{ij} =length of time in the jth interval of the flow measurement in the ith interval of the concentration measurement (T). This mid-discharge splitting method can also be applied for measurements at unequal time-interval bases. Alternatively, if the concentration measurements are based on constant discharge volume, the weighted average of $w_i c_i$ reduces to the arithmetic average. Ideally, automated samplers can collect samples in proportion to discharge volume.

The MFF ratio is defined as the ratio of the normalized discharge mass of pollutants to the normalized runoff volume in the first portion of the total runoff volume and mathematically expressed as follows (Ma et al. 2002; Kim et al. 2004):

$$\mathrm{MFF}_n = \frac{\int_0^{t_n} C(t)Q(t)dt \bigg/ M}{\int_0^{t_n} Q(t)dt \bigg/ V} = \frac{\int_0^{t_n} C(t)Q(t)dt \bigg/ \int_0^{t_n} Q(t)dt}{M/V} \tag{6}$$

where MFF_n=mass first flush ratio for the first n% of total runoff volume; t_n=elapsed time corresponding to the first n% of total runoff volume. In this study, the MFF_n for 20% of total runoff volume (MFF_{20}) was used for the first flush effect simulation. To calculate MFF_{20} using Eq. (6), continuous functions of pollutant mass discharge and pollutant concentrations are needed, and the same techniques used to develop approximate continuous functions for the EMC were used.

Simulations of Different Sampling Strategies

Reference Value for EMC

To evaluate different sampling strategies, a best estimate or reference value of pollutant concentration is needed for comparison. The reference value will be used when simulating various sampling strategies. The reference pollutant concentrations for a

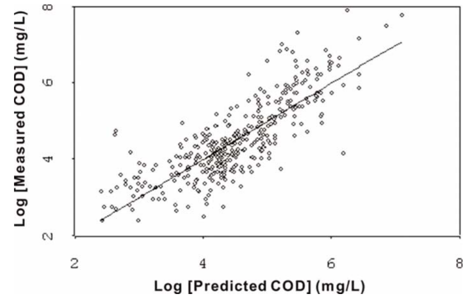

Fig. 2. Regression's fitted values versus observations

storm event were determined using a previously developed COD regression model (Ma 2002), developed using regression analyses with 393 grab sample COD concentrations from the three highway sites during the 1999–2001 wet seasons. The mathematical expression of the model is as follows:

$$E(\log \mathrm{COD}_i) = 6.08 - 0.60 \log \mathrm{CumRF}_i + 0.40 \log \mathrm{ADD} - 0.16 \log \mathrm{AntRF} \tag{7}$$

$$\varepsilon_i \sim N(0, 0.59^2) \tag{8}$$

where $E(\log \mathrm{COD}_i)$=mean value of log transformed chemical oxygen demand concentration given independent variables for ith grab sample (mg/L); CumRF_i=cumulative rainfall at the time of the ith grab sample (1 mm increments); ADD=antecedent dry days before the monitored event (days); AntRF=previous event's precipitation before the monitored event (1 mm increments); and ε_i=error of prediction having normal distribution. Only a very short lag between rainfall and runoff was observed and the impact of time of concentration has been discussed elsewhere (Kang et al. 2008). Fig. 2 shows the model's fitted values versus the observations. Eqs. (7) and (8) were used to generate COD concentrations at 1-min intervals, which is the shortest possible sampling frequency, since the rainfall and flow data are collected at 1-min intervals. Since the regression model predicts mean values of log-transformed COD, which is median values in arithmetic scale distribution, the mean value of COD is obtained using the relationship of lognormal distribution (Aitchison and Brown 1957) as follows:

$$E(\mathrm{COD}_i) = e^{E(\log \mathrm{COD}_i) + \sigma^2/2} \tag{9}$$

where $E(\mathrm{COD}_i)$=mean value of COD given independent variables for ith grab sample; and σ^2=variance of prediction errors (=0.59^2). The EMC and MFF_{20} were then calculated using Eqs. (2) and (7), respectively, and used as the reference values of EMC or MFF_{20}. The EMCs and MFF_{20}s generated represent the most precise estimates, as if a grab sample for COD could be collected and analyzed each minute.

Simulations of Sampling Strategies

In order to simulate the collection of a varying number of grab samples, Eq. (7) was used with a random component (white noise) having mean zero, and a variance equal to the variance in the original data [Eq. (8)]. In this way, collecting any number of grab samples with measurement errors can be simulated. To compare different sampling strategies, five types of simulation were performed using different numbers and different strategies for collecting samples during typical storm events as described in

Table 1. Simulations of Different Sampling Strategies

Simulation type		Sample size (n) or sampling interval	Sampling type
1	Randomly timed	$n=10$, 20, 40, 60, 100, and 1-min interval	Manual
2	Equally timed	$n=10$, 20, 40, 60, 100, and 1-min interval	Manual
3	Equal-rainfall depth	$n=10$, 20, 40, 60, 100, and 1-min interval	Automatic, manual
4	Equal-discharge volume	$n=10$, 20, 40, 60, 100, and 1-min interval	Automatic, manual
5	First flush enhanced	First flush sampling: • 15-min interval for first 1 h • 30-min interval for second 2 h • 1-h interval for the rest of the storm Equal-time sampling: • 45-min interval for whole storm	Manual

Table 1. A total of 35 different rainfall patterns, corresponding to actual observed patterns in our monitoring program, were used. The hydrographs were smoothed before simulation to correct short-term fluctuations in the original data. Table 2 is a statistical description of the 35 events.

In order to estimate the variability that might occur when using white noise to simulate storm variability, each type of simulation was performed 1,000 times to generate a mean value and distribution of the EMCs and MFF_{20}s. It was expected that strategies that collected more grab samples will produce the EMCs or MFF_{20}s that are closer to the reference values. By comparing the number or timing of grab sample collection, each strategy can be compared. The influence of sample size on calculated EMCs or MFF_{20}s was evaluated by simulating 10, 20, 40, 60, and 100 samples per event using five timing strategies, which will be referred to as Types 1 to 5, described in Table 1 and the following paragraphs. These strategies were compared to ideal or reference sampling, where samples are collected every minute. These strategies all examine errors in concentration measurements and ignore potential errors in flow measurements. Measurements in flow can also occur, but their impacts were not considered in this work, and is a subject for future work.

Type 1 uses randomly timed grab sampling. The required number of samples (n) is collected at random points of time during each storm event. Theoretically, this is the most general case for a sample set with fixed size. Type 2 uses equally timed or spaced sampling. The required number (n) of grab samples is collected at equal time intervals during the runoff. To avoid unique results that might occur because of a specific sample sequence, each sample sequence is shifted forward or backward based upon a random time, ranging from -5 to $+5$ min. Type 3 simulations used equal rainfall depth sampling. A specified number of samples are collected at equal intervals of rainfall depths. Type 4 simulations used equal discharge volume sampling. A specified number of samples are collected at each point of time when an equal runoff volume is discharged. Any error in flow measurement error is ignored. Type 3 and 4 strategies usually require automated samplers, and Type 4 is often used to collect flow-weighted composite samples, as shown in Fig. 1. Type 5 was a simulation of more frequent sampling in the earlier part of a storm in order to improve the characterization of the first flush. In the first flush enhanced sampling, five samples are taken in the first hour. The first sample is taken at time zero and the remaining samples are taken at 15-min intervals. Next, four samples are taken in the following 2 h at 30-min intervals and afterward, one sample is taken every 1 h. The results of all sampling types were compared with each other and to the reference sampling.

Fig. 3 illustrates the results of Type 2 sampling (equally timed), showing the frequency distribution of 1,000 simulated EMCs using different sample sizes for one event (1/25/2,000, site 1, total rainfall depth=17 mm, storm duration=19.4 h). The 1-min sampling simulation was used for producing a reference value of true COD measurement per storm event (EMC^R). Moreover, the 1-min sampling was not just simulated once but 1,000 times for each storm event to generate a sample set consisting of 1,000 reference values and the average of them was used as EMC^R. For the same storm event under less frequent sampling strategies (i.e., Type 2 sampling as $n=5$, 10, 20, 40, and 60 as illustrated in Fig. 3), each sampling simulation generates a set of EMCs from the 1,000 runs. The error of each sampling strategy is estimated by comparing the distribution of the EMCs generated from 1,000 runs to the EMC^R obtained from 1-min simulation. Specifically it is estimated by the root-mean-square error (RMSE) that analyzes the bias of sample mean as well as the variance of target distribution. The RMSE result was then normalized by converting into percentage (mathematical expression is shown in the next section). In Fig. 3, EMC^R is 116.26 (mg/L), and the means of each of the 1,000 simulations for $n=10$, 20, 40, 60, and 1-min interval samples are 149.15, 121.48, 117.91, 117.12, and 116.22 (mg/L), respectively. The results indicate that smaller sample size produces larger deviation from the true EMC, which is expected since the COD is log normally distributed.

Table 2. Hydrologic Characteristics for 35 Monitored Events

Hydrologic property	Average	Standard deviation	Minimum	Median	Maximum
Total rainfall (mm)	29.7	39.1	2.0	17.0	156.0
Maximum rain intensity (mm/h)	7.9	8.4	0.5	4.8	32.5
Discharge volume (m^3)	285	377	7	140	1,422
Maximum discharge rate (m^3/min)	1.3	1.2	0.06	1.0	5.6
Rain duration (min)	661	513	93	610	2,376

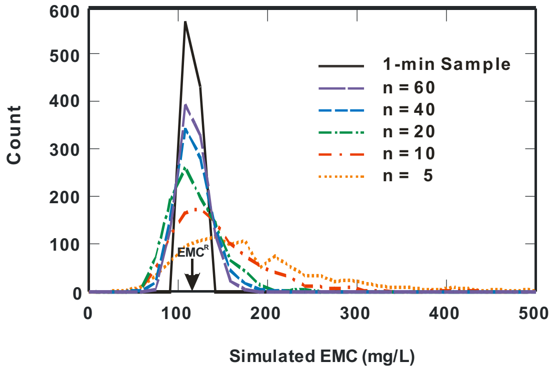

Fig. 3. Frequency distribution of EMCs from different sample numbers for the Type 2 simulations (storm event recorded on 01/25/2000, Site 1, total rainfall depth=17 mm, storm duration=19.4 h

Percentage Error Calculation for EMCs

To quantify the average error between the simulated EMC and EMC^R, the RMSE was used. The RMSE is equal to the bias and the variance of the estimate, noted by EMC* and simply calculated using Eq. (10)

$$\text{RMSE}(\text{EMC}^*) = \sqrt{\Sigma_{i=1}^{N}(\text{EMC}_i^* - \text{EMC}^R)^2 / N}$$
$$= \sqrt{(\overline{\text{EMC}^*} - \text{EMC}^R)^2 + \text{Var}(\text{EMC}^*)} \quad (10)$$

where N=number of runs (=1,000), EMC^*=EMC from a single simulation and $\overline{\text{EMC}^*}$=mean of EMC^* over 1,000 runs and EMC^R=reference EMC. The percentage error (Err%), based on RMSE, is used to quantify the estimation errors as

$$\text{Err}\,\% = \frac{\text{RMSE}(\text{EMC}^*)}{\text{EMC}^R} \times 100\% \quad (11)$$

The Err% for the 1-min sampling is caused only by its variance. The Err% of the 1-min sample EMC for the case in Fig. 3 is 3%. It should be noted that all the Err% calculations were focused on the prediction errors originating from only sampling (flow measurement errors were not considered) by using the same flow data for both the sampling simulations and the reference value calculations.

Results and Discussion

Estimation Errors in EMC

The calculated values of Err% from the various simulations for different types of sampling strategies are presented in a series of box plot figures, showing medians and 50% interquartile ranges with outliers. Fig. 4 shows the Err% distribution of simulated EMCs for varying numbers of samples. Fig. 4(a) shows the results from Type 1 simulation. The largest Err% was approximately 80% at n=10. The mean values of Err% at n=10, 20, 40, 60, and 100 were 46.2, 28.4, 19.3, 15.4, and 11.9%, respectively. The median values of Err% were slightly lower than the average values. The corresponding standard deviations were 12.5, 6.1, 4.0, 3.3, and 2.3%. The Type 1 simulation can serve as a benchmark on the influence of sample size for estimating the EMCs, and is the most general sampling strategy. Random sampling inherently produced large errors in the estimated EMCs and, therefore, required larger than 40 samples for 75% of the EMC estimates to have less than 20% error; 60 samples estimate the EMCs within 20% error in most cases.

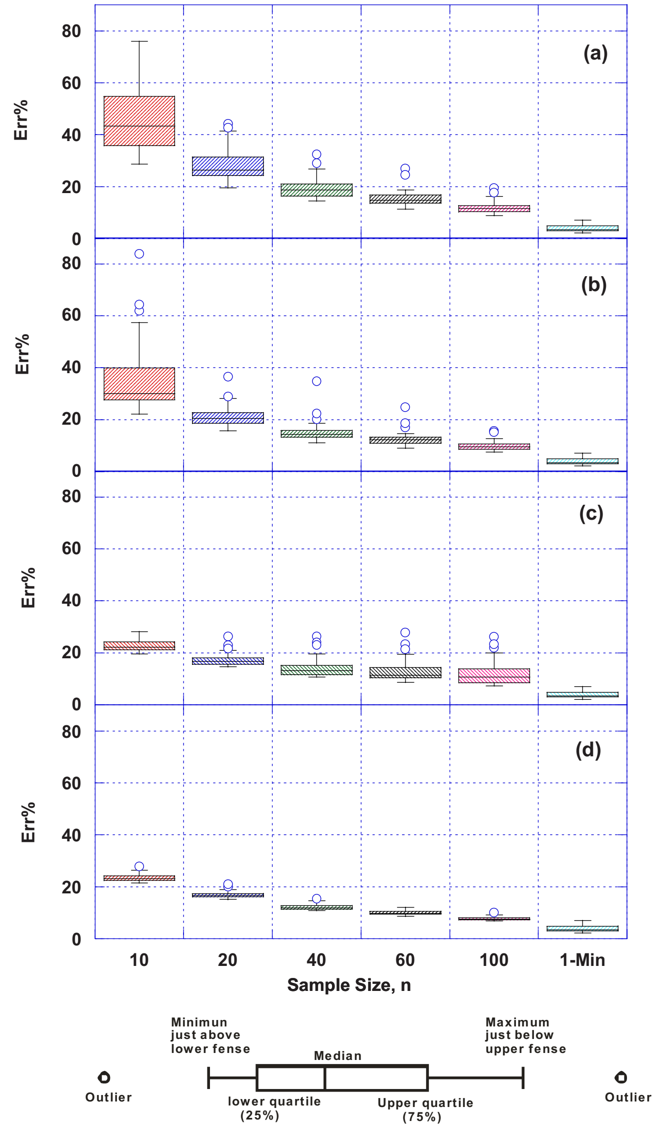

Fig. 4. Percentage errors in the EMCs calculated with different sample sizes for different simulation types with a prediction error of $N(0,0.59^2)$: (a) randomly timed sampling (Type 1); (b) equally timed sampling (Type 2); (c) equal-rainfall depth sampling (Type 3); and (d) equal-discharge sampling (Type 4)

Fig. 4(b) shows the Err% distributions from the Type 2 simulations. The worst case was n=10 with Err% less than 60% except three points of outliers, which reflects an improvement over Type 1 simulation (randomly timed sampling). The mean values of Err% at n=10, 20, 40, 60, and 100 were 36.3, 21.3, 15.2, 12.6, and 9.8%, respectively. The median values of Err% were generally less than the mean values. The corresponding standard deviations were 13.5, 4.1, 4.2, 2.9, and 1.9%. These statistics imply an improvement over randomly timed sampling. When this sampling strategy is used, approximately 40 samples are required to estimate the EMCs within 20% error.

Fig. 4(c) shows the Err% distributions from Type 3 simulation. At n=10, the worst case is approximately 30%, which reflects a large improvement over Type 1 or Type 2 simulation. The mean values of Err% at n=10, 20, 40, 60, and 100 were 22.9, 17.5,

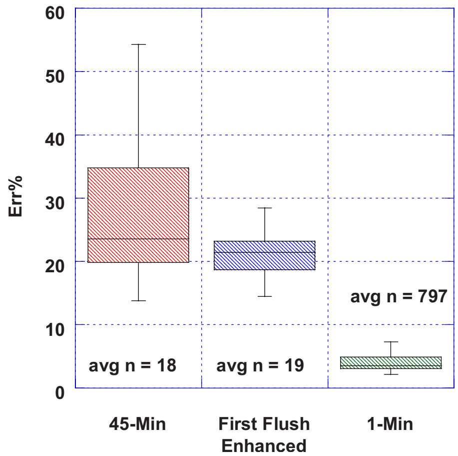

Fig. 5. Comparison of percentage errors in the EMCs from equal-time sampling and first flush enhanced sampling simulations (Type 5)

14.2, 13.1, and 12.2%, respectively. The median values of Err% were generally the same as the mean values. The corresponding standard deviations were 2.2, 2.6, 3.8, 4.4 and 4.8%. As shown in Fig. 4(c), the interquartile range increases with an increase in the number of samples for Type 3, which contradicts intuition. This results due to the stepwise nature of the recorded cumulative rainfall depth. The data logger only receives a signal when a unit amount of rainfall depth (0.01 in) is accumulated in the tipping bucket of the rainfall gauge. As the sample number increases for a given storm duration, the time interval between two consecutive sampling points becomes smaller than the time required to detect the increment of rainfall depth, resulting in duplicate sampling points in the simulation and, thereby, increasing variation among different storm events that have different rainfall characteristics (intensity and duration). This sampling strategy was better than random or equal-time sampling in terms of accuracy in EMC estimation. Less than 20 samples were required for 75% of the EMC estimates to be within 20% error.

Fig. 4(d) shows the Err% distributions from the Type 4 simulations. It is obvious from the plots that Type 4 simulation has the best result from the aspect of outliers, averages, and variances. The mean values of Err% at $n=10$, 20, 40, 60, and 100 were 23.4, 17.0, 12.2, 10.0, and 7.8%, respectively. The median values are generally the same as the mean values. The corresponding standard deviations are 1.4, 1.3, 1.1, 0.9 and 0.7%. Compared to equal rainfall interval sampling, equal discharge volume sampling estimates the EMCs more accurately and precisely. The EMCs were estimated with less than 20% error in most cases using only 20 samples. Approximately twice the number of samples was required for similar accuracy using equally timed sampling, and three times as many samples were required for randomly timed samples.

Fig. 5 shows the Err% distributions from the Type 5 simulation. The average numbers of samples taken are 19 for first flush enhanced sampling and 18 for the 45-min interval sampling (equally timed sampling). The corresponding means and standard deviations of Err% are 21.3 and 3.8% for first flush enhanced sampling, and 27 and 9.9% for 45-min interval sampling, respectively. Although the average sample numbers of both methods differs only by one sample, the accuracy of the estimated EMC was greatly improved using the first flush enhanced sampling method. This occurs because the more frequent sampling at the beginning of the storm is able to capture the rapid decline in pollutant concentration (Li et al. 2005).

Accuracy in MFF Estimation

Fig. 6 shows the values of MFF_{20} with respect to values of MFF_{20} obtained from Type 2 [Figs. 6(a and b)], Type 4 [Figs. 6(c and d)], Type 5 [Fig. 6(e)], and 1-min sampling [Fig. 6(f)] simulations. Each data point with vertical error bars indicates the median with 25 and 75% quartiles of 1,000 runs for each storm event simulated. The values of R^2 shown in the figure were calculated using the medians and the line of equivalence. A larger number of samples routinely produced more accurate MFF_{20} with less statistical dispersion of random trials (distance between 25 and 75% quartiles).

As shown in Figs. 6(a) and Fig. 6(b), equally timed sampling tends to overestimate (most medians lie above the equivalence line) the MFF_{20} values with larger deviations from reference values for fewer numbers of samples. This is because the limited numbers of grab samples cannot capture the rapid decline in COD concentration in the initial runoff. In contrast, Figs. 6(c and d) show that equal-discharge sampling tends to underestimate (all of median values lie under the equivalence line) the MFF_{20} values. In the equal-discharge sampling strategy, initial sampling occurs after observing a specified amount of runoff, and, therefore, the delay of sampling initiation, especially in low initial runoff rate, misses the high concentration of the initial runoff, underestimating mass first flush. A larger number of samples reduced this underestimation problem (values of R^2 at $n=20$ and $n=40$ are 0.39 and 0.84, respectively), and the variation in MFF_{20} values from 1,000 runs. When an automatic sampler is used, a shorter sampling pace will provide better MFF estimation.

By comparing the results from Type 2 and Type 4 simulations with same numbers of samples in Fig. 6 [i.e., Figs. 6(a and c) or, Figs. 6(b and d)], one can observe that equally timed sampling can estimate the MFF_{20} more accurately than equal-discharge sampling when a relatively small number of sample size is used (values of R^2 for equally timed and equal discharge sampling at $n=20$ are 0.84 and 0.39, respectively). The opposite results occurred for EMC estimation; equal-discharge sampling is always superior to equally timed sampling for the EMC estimation, regardless of the number of samples. This can also be explained because the equal-discharge sampling collects fewer samples in the early runoff, when the flow is still increasing. The first flush enhanced sampling strategy overcomes this problem; Fig. 6(e) shows better estimates compared to Fig. 6(c) for almost the same number of samples.

Potential for Improving Sampling Procedures

Effect of Prediction Error

It is burdensome and expensive to collect so many grab samples, and it is natural to ask what improvements can be made to reduce this burden. If the prediction error (ε_i) could be reduced, perhaps by using an alternative analysis with less intrinsic variability, how could the number of samples be reduced? According to the simulation results, if the error were reduced by 50% (standard deviation$=$0.30), the means and medians of Err% can also be reduced by about 50% in Type 2, Type 3, and Type 4 sampling simulations. When using a small number of samples ($n=10$, 20) in the Type 1 simulation, a reduced amount of the prediction error

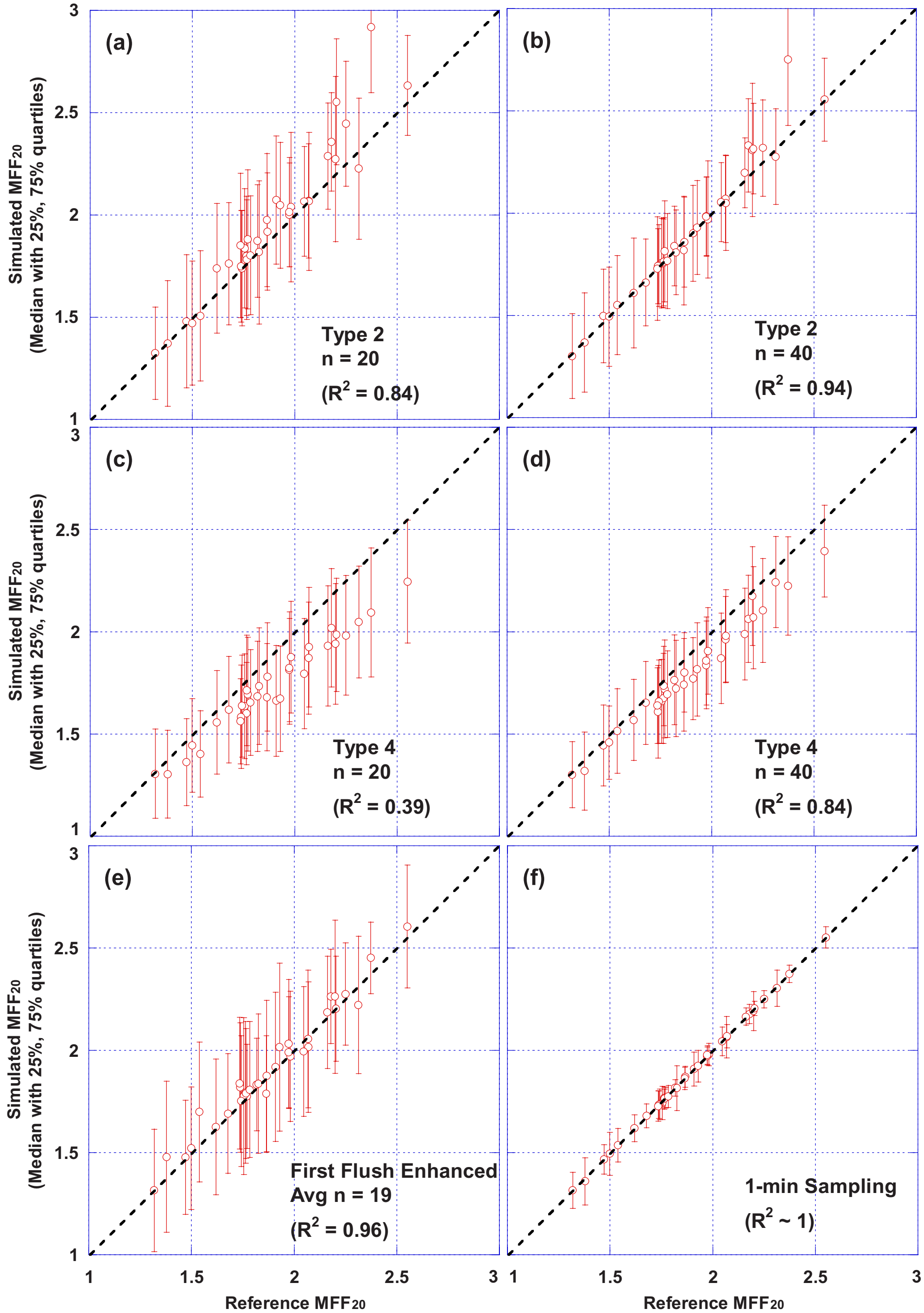

Fig. 6. Comparison between simulated and reference values of MFF_{20} for different sampling strategies (values of R^2 were calculated using median values and equivalent line)

could not proportionally reduce the Err% estimation. Fig. 7 shows the Err% distribution of simulated EMCs for varying numbers of samples with 0.3 of σ for different sampling strategies. Twenty samples provide less than 20% errors in 75% of the cases even in worst sampling strategy (i.e., Type 1 randomly timed sampling). Most of the cases had less than about 15% of Err% in Type 2, Type 3, and Type 4 sampling simulations with $n=20$. When using Type 3 or Type 4 simulation, only 10 samples provide less than 10% of Err% in most cases and larger than 40 samples did not provide a dramatic improvement in EMC estimation (note that larger than 20 samples were required for less than 20% of Err% at $\sigma=0.59$). This indicates that in order to obtain accurate EMCs, reducing the prediction error can also be important.

Extended Sampling

Another alternative is to collect extended grab samples, which can be collected over a short period of time to composite the concentrations. Extended sampling can average out the sampling

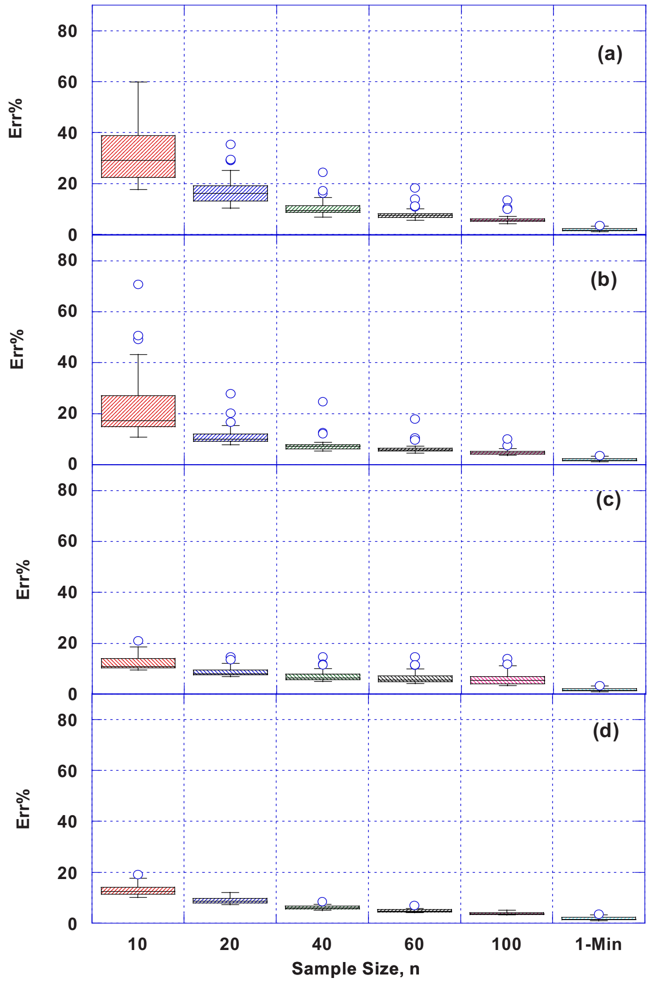

Fig. 7. Percentage errors in the EMCs calculated with different sample sizes for different simulation types with a prediction error of $N(0,0.30^2)$: (a) Type 1 (randomly timed sampling); (b) Type 2 (equally timed sampling); (c) Type 3 (equal-rainfall depth sampling); and (d) Type 4 (equal-discharge sampling)

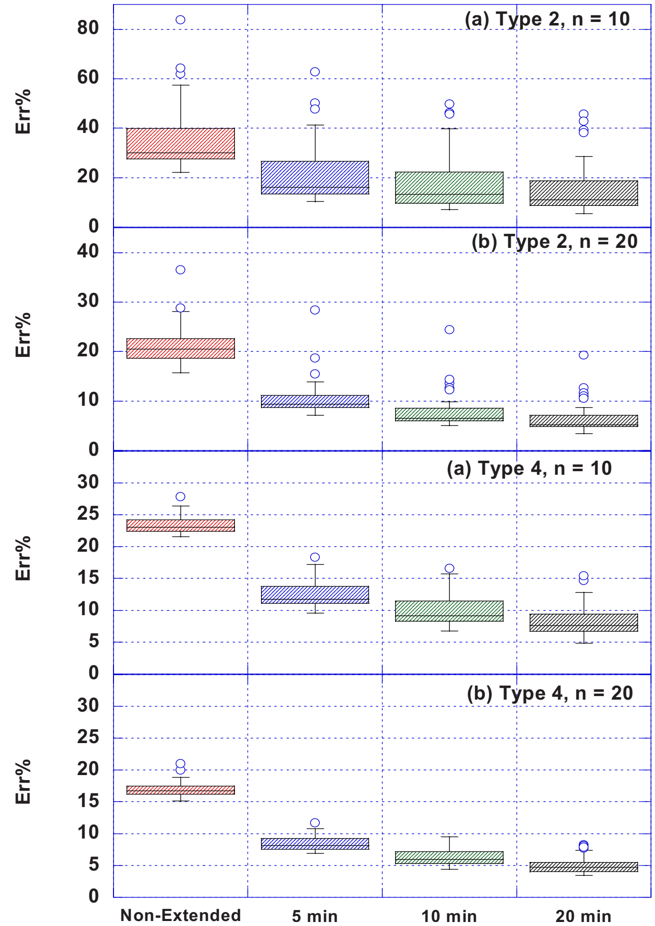

Fig. 8. Percentage errors in the EMCs calculated with different sampling time extension in Type 2 and Type 4 sampling strategies with a prediction error of $N(0,0.59^2)$: (a) Type 2, $n=10$; (b) Type 2, $n=20$; (c) Type 4, $n=10$; and (d) Type 4, $n=20$

error to approach zero, reducing the overall prediction error. In order to simulate extended sampling, each grab sample was assumed to be a composite of 1-min interval samples over a specified extended sampling time (5, 10, and 20 min). Fig. 8 shows the effect of the extended sampling in reducing Err% of simulated EMCs for different extended times and sample numbers ($n=10,20$) in the Type 2 and Type 4 sampling strategies. When an extended composite sample was collected over 5 min in 1-min intervals, the median Err% was reduced by approximately 50% for all cases. Extended sampling can also improve the MFF estimation in terms of medians and dispersions as shown in Fig. 9. Fig. 9 compares the MFF_{20} estimation using 20-min extended sampling at $n=20$ using the Type 2 and Type 4 sampling strategies. Compared to the regular sampling [in Figs. 6(a and b)], extended sampling significantly reduces the dispersion of simulated MFF_{20}. As shown in Figs. 6 and 9, extended sampling could be more efficient for the Type 4 sampling than for the Type 2 sampling. The collection of an extended grab samples over a short time would not be so burdensome to the sampling program. The field time for crews would be only slightly extended, and the transportation and analysis would be the same, because the grab samples collected during the extended time are composited into a single bottle for analysis.

Conclusions

This paper has shown that a flow-weighted composite sample can be viewed as a series of grab samples summed with weights that reflect the flow. To evaluate the error of using a limited number of grab samples and the strategy for collecting the samples, a series of simulations was performed using a COD correlation, random noise, and hydrographs from 35 different storm events.

The results showed that a series of 10 flow-weighted grab samples provides a relatively poor estimate of the EMC, with median RMSEs of 46% for randomly timed samples to 23% for samples collected at equal discharge volumes. If the number of grab samples increases to 20, the median error is reduced to 28% for randomly timed samples to 17% for samples collected at equal discharge volumes. Even if 100 samples are collected, the error is still nearly twice as large as the minimum possible error, when samples are collected each minute.

The best strategy for EMC estimation is to collect the grab samples at equal discharge volume intervals. Equal rainfall interval is the second choice, with equal timing and random timing

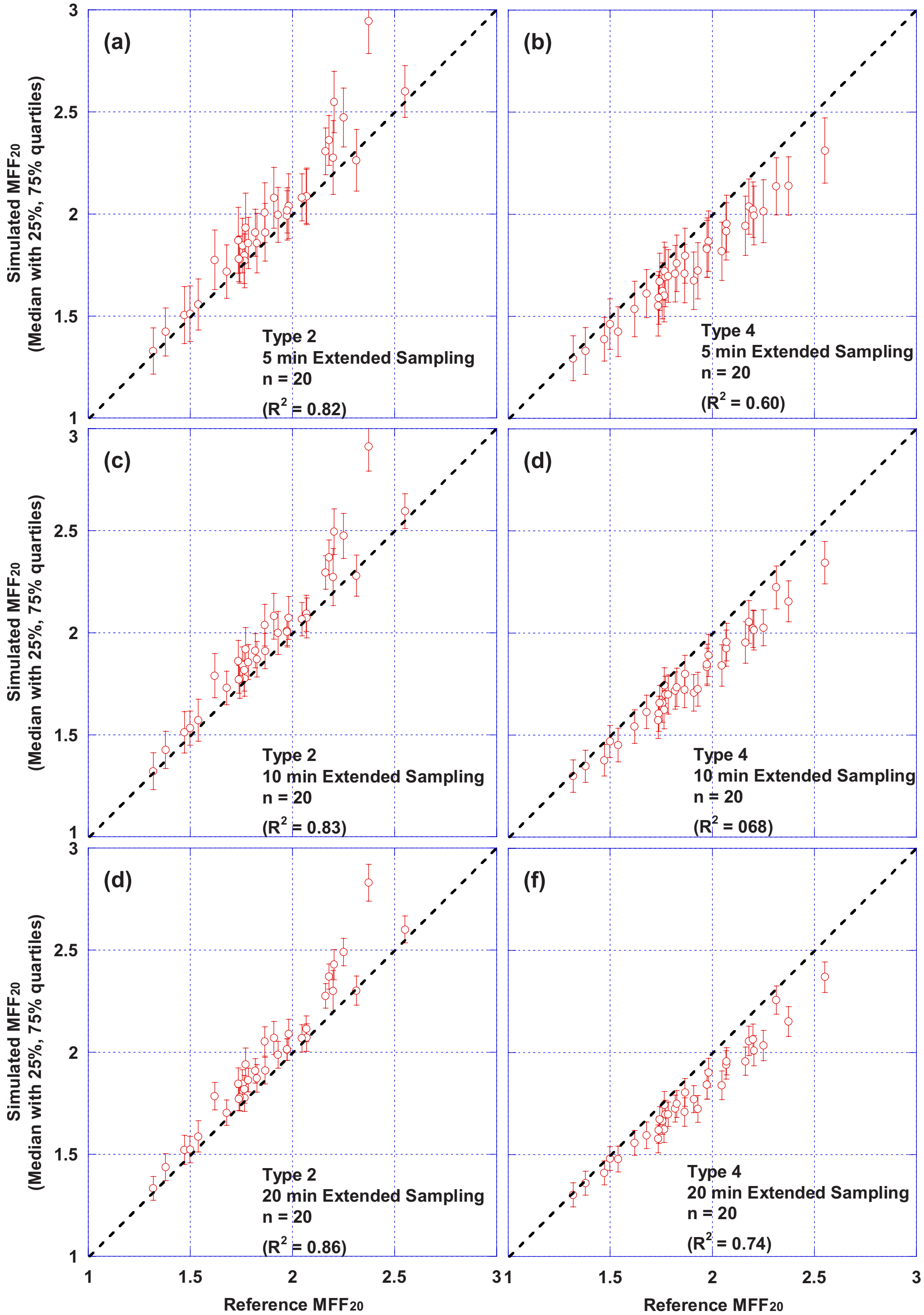

Fig. 9. Comparison between simulated and reference values of MFF_{20} using 20-min extended grab sampling time for Type 2 and Type 4 sampling strategies

being less desirable strategies. Reducing the sampling error can greatly improve the accuracy in EMC estimation. However, equal discharge grab sample with relatively large sampling pace might not be a good strategy for MFF estimation because the small number of samples will underestimate mass discharge rate of the initial runoff. For MFF estimation, an automatic sampler with short sampling pace is the best choice and larger sample numbers or first flush enhanced sampling will be next best.

The results show that automatic flow weighted samplers, which can be programmed to collect several hundred subsamples per storm, are far superior to collection by grab sampling, even if 100 grab samples are used. If automatic samplers can be used without chemical or physical biases (e.g., such as the concerns of sample carry over when sampling for oil and grease, or the introduction of artifactual toxicity), they should be used. A new concept, extended grab sampling, which uses several grab samples collected over a short time is a promising way to reduce sampling variability when composite samples are not practicable.

Acknowledgments

This project was partially supported by the Division of Environmental Analysis, California Dept. of Transportation through a contract with the University of California and task order No. 43A0073. The writers appreciate the help of Dr. Simlin Lau, who performed the analytical measurements.

References

Aitchison, J., and Brown, J. A. C. (1957). *The lognormal distribution*, Cambridge University Press, Cambridge, U.K.

Carleton, J. N., Grizzard, T. J., Godrej, A. N., and Post, H. E. (2001). "Factors affecting the performance of stormwater treatment wetlands." *Water Res.*, 35(6), 1552–1562.

Charbeneau, R. J., and Barrett, M. E. (1998). "Evaluation of methods for estimating stormwater pollutant loads." *Water Environ. Res.*, 70(7), 1295–1302.

Gulliver, J. S., and Anderson, J. L. (2008). *Assessment of stormwater best management practices*, Univ. of Minnesota, Minneapolis.

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

Kang, J. H., Kayhanian, M., and Stenstrom, M. K. (2008). "Predicting the existence of stormwater first flush from the time of concentration." *Water Res.*, 42(1–2), 220–228.

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

Kim, L.-H., Kayhanian, M., Zoh, K.-D., and Stenstrom, M. K. (2004). "Modeling of highway stormwater runoff." *Sci. Total Environ.*, 348(1–3), 1–18.

Li, Y., Lau, S.-L., Kayhanian, M., and Stenstrom, M. K. (2005). "Particle size distribution in highway runoff." *J. Environ. Eng.*, 131(9), 1267–1276.

Ma, J. S. (2002). "Characteristics of pollutants in highway runoff: Regression, representativeness, and first flush." Ph.D. thesis, Univ. of California, Los Angeles, Los Angeles.

Ma, J.-S., et al. (2002). "First flush phenomena for highways: How it can be meaningfully defined." *Proc., 9th Int. Conf. on Urban Drainage*, ASCE, Reston, Va., 112, 223.

Stenstrom, M. K., and Kayhanian, M. (2005). "First flush phenomenon characterization." *CTSW-RT-05-73-02.6*, Calif. Dept. of Transportation, Division of Environmental Analysis, Sacramento, Calif.

United States Environmental Protection Agency (USEPA). (1983). "Results of the nationwide urban runoff program." *Vol. I—Final Rep.*, Water Planning Division, WH-554, Washington, D.C.

United States Environmental Protection Agency (USEPA). (2002). "Urban stormwater BMP performance monitoring." *821-B-02-001*, Office of Water, Washington, D.C.

Wu, J. S., Allan, C. J., Saunders, W. L., and Evett, J. B. (1998). "Characterization and pollutant loading estimation for highway runoff." *J. Environ. Eng.*, 124(7), 584–592.