Sampling for PPCPs in Wastewater Systems: Comparison of Different Sampling Modes and Optimization Strategies

CHRISTOPH ORT,*,† MICHAEL G. LAWRENCE,† JULIEN REUNGOAT,† AND JOCHEN F. MUELLER‡

The University of Queensland, Advanced Water Management Centre (AWMC), QLD 4072, Australia, and The University of Queensland, National Research Centre for Environmental Toxicology (EnTox), Coopers Plains, QLD 4108, Australia

Received March 10, 2010. Revised manuscript received May 26, 2010. Accepted June 17, 2010.

The aim of this study was to assess uncertainties associated with different sampling modes when evaluating loads of pharmaceuticals and personal care products (PPCPs) in sewers and influents to sewage treatment plants (STPs). The study demonstrates that sampling uncertainty can range from "not significant" to "far greater than the uncertainty due to chemical analysis", which is site- and compound-specific and depends on the (in)accuracy of the analytical method. Conventional sampling devices operated in common time- or flow-proportional sampling modes, and applying traditional sampling intervals of 30 min or longer can result in the collection of nonrepresentative samples. At the influent of a STP, wastewater may appear as a continuous stream, but it is actually composed of a number of intermittently discharged, individual wastewater packets from household appliances, industries, or subcatchments in pressurized sewer systems. The resulting heterogeneity can cause significant short-term variations of pollutant loads. We present different experimental results and a modeling approach showing that the magnitude of sampling uncertainty depends mainly on the number of pollutant peaks and the sampling frequency; sampling intervals of 5 min or shorter may be required to properly account for temporal PPCP variations in influents of STPs. A representative sample is a prerequisite for providing meaningful analytical results and cannot be compensated with a large number of samples, accurate chemical analysis, or sophisticated statistical evaluation. This study highlights that generalizing from one case to another is difficult and hence a careful systems analysis of the catchment under investigation, or precautionary choice for a sophisticated sampling mode, is necessary to prove reproducibility.

Introduction

Increasingly, raw wastewater samples are analyzed for pharmaceuticals and personal care products (PPCPs). Wastewaters are highly variable streams and the daily volumes are large. Flow volumes can range from 100 $m^3\ d^{-1}$ (e.g., effluent of a regional hospital or a small municipal subcatchment) up to 100,000 $m^3\ d^{-1}$ or even more (i.e., influent to large sewage treatment plants (STPs)). The lack of online instrumentation, the unfavorable physical and chemical properties of wastewater, and the harsh conditions in sewers do not (yet) allow for a direct analysis of PPCPs. Therefore, samples have to be collected to quantify PPCP concentrations in a laboratory.

Automated sampling devices are used to collect a number of discrete samples over a certain period, usually 24 h. Three facts result in relatively long sampling intervals when using commercially available sampling devices: (i) the minimum sample volume for an individual discrete sample, (ii) the limited refrigerated storage capacity, and (iii) the reserve to account for high flows (rain events) to avoid overflowing or stopping of the sampling device. However, the choice of an appropriate sampling interval should be informed by the expected concentration variability, and not by the limitations of the sampling device.

To minimize analytical costs, the discrete samples are pooled to a composite sample, typically already as a result of the sampling process (collection in one bottle). The result is a manageable volume of 1−20 L; this volume is then subsampled to determine that sample's *average* concentration in the laboratory. The volume of this subsample generally ranges from 1 mL (for direct injection) to several hundred mL (for solid phase extraction).

Due to the pooling step and the absence of suitable online instrumentation, data on site-specific short-term fluctuations of PPCPs in sewers are, to a large extent, lacking. Based on recommendations in internationally accepted water quality monitoring norms the opposite would be expected: "The times and frequencies of sampling in any programme can be properly decided only after detailed preliminary work, in which a high sampling frequency is necessary [...]" ISO, 1980 (*1*). Maybe preliminary investigations are simply not reported or not recognized to be sufficiently important (*2*). The high analytical costs per sample appear to be another reason: preliminary work at one location and one point in time may not be transferable and would have to be repeated, adding substantial costs to monitoring campaigns without answering the final research questions.

To our knowledge, only two studies have presented time series of micropollutants at high temporal resolution in sewers; and both focus only on benzotriazole contained in dishwashing detergents (*3, 4*). The measured time series proved to show properties similar to those of selected realizations from a stochastic, predictive model that was set up to characterize pollutant patterns in sewers (*3*). Subsequently, the model was used to investigate how the combination of PPCP short-term fluctuations and sampling frequencies affect the representativeness of composite samples. The results revealed that the evaluation of PPCP loads in sewers can be prone to significant uncertainty. Consequently, the application of the modeling approach was suggested to overcome the constraints of high-frequency sampling campaigns to determine the appropriate sampling frequency (*4*).

Relevant sampling guidelines have existed for decades (*1, 5–7*), yet a review of 87 papers comprising 267 different sewer sites reveals that these published procedures and methods are neither cited nor heeded (*2*). As a result, it is not understood how accurately the concentration of a chemical in the influent of a STP is represented by a very small sample volume (i.e., a fraction as small as 10^{-11} of the large wastewater

* Corresponding author e-mail: c.ort@awmc.uq.edu.au; tel: +61 (0)7 3365 4730; fax: +61 (0)7 3365 4726.
† AWMC.
‡ EnTox.

10.1021/es100778d
Published on Web 07/15/2010

stream). Therefore, we present three *approaches* to address three key aims in this study:

1. High-Frequency Grab Sampling. Distinct toilet flushes and wastewater packets from other household appliances are expected to determine the highly variable PPCP patterns in (gravity) sewers. It shall be tested whether individual wastewater packets can be identified in a continuous wastewater stream; if yes, their temporal extension shall be determined.

2. Different Composite Sampling Modes. Relatively long sampling intervals may not be suitable to representatively sample for PPCPs in sewers given the unknown and potentially high short-term concentration variations. The aim of this part is to provide sound experimental evidence that commonly applied sampling modes and frequencies can lead to significant sampling artifacts that can potentially overwhelm analytical uncertainty.

3. Modeling Study. An existing model is extended to determine an appropriate sampling mode and frequency in the planning phase, or to retrospectively assess the sampling uncertainty of past sampling campaigns.

Our results highlight an aspect of sampling uncertainty that has not previously been comprehensively addressed in the context of PPCPs in sewers and influents of STPs (raw wastewater), and clearly demonstrate that sampling uncertainty can be a dominant source of error. Are these results relevant to your studies? Given the large variation of unique catchment properties (sewer systems), the vast number of micropollutants, the various specific sampling modes and distinct analytical methods, there are numerous possible combinations. The presented examples cannot cover the entire range, but emphasize the need for a careful evaluation of each individual case. To rule out uncertainties related to sampling for particulate matter we focus on quantifying compounds in the dissolved phase to investigate the effect of temporal (short-term) variations on sampling. Furthermore, in most studies only dissolved concentrations are measured and reported (*8*).

Material and Methods

Description of Sewer Catchments. Sewage treatment plant A is gravity-fed and receives the wastewater from approximately 100,000 inhabitants connected to a separate sewer system (negligible surface runoff); the wastewater is not pumped into STP A. In contrast, STP B (total number of inhabitants approximately 45,000) receives the wastewater from a combined sewer system with a total of 71 sewage pumping stations. One third of the catchment is mainly pressurized and the wastewater is stored at certain times of the day in a large retention tank. The other part ($^2/_3$) is mainly gravity sewers. Nonetheless, all wastewater needs to be lifted at the influent of STP B. The special configuration of the inlet work at STP B implies high flow variations, yet this configuration is expected to *reduce* the magnitude of short-term concentration variations. More detailed information, including typical diurnal variations (flow patterns), is presented in Table SI 1 (Supporting Information).

High-Frequency Grab Sampling. Anthropogenic gadolinium (Gd_{anth}), the Gd used as contrast agent for magnetic resonance imaging (MRI) (*9*), was selected as the target compound for the high-frequency grab sampling campaign at STP A. The two main reasons were as follows: (i) The application of Gd_{anth} is limited to the use as MRI agent and even in a catchment of a large STP, only a few toilet flushes containing Gd_{anth} are expected. Previous long-term investigations in the *effluent* of STP A indicated that daily Gd_{anth} loads originated from the treatment of one to two patients with the highest loads of contrast agents corresponding to midweek applications (*10*). The elimination half-life of contrast agents in the human body is 1.3−2 h. Thus, it could be expected that >99.9% of a full dose would be discharged to the sewer over the typical five toilet flushes within 24 h (*11*). (ii) Gadolinium can be quantified by ICP-MS at the 10^{-15} mol L^{-1} level (*12*). Maximum influent concentrations are expected to be up to six orders of magnitude higher in the most concentrated samples which will depend on the instantaneous dilution of a toilet flush in the total wastewater stream. With an instrumental uncertainty of <0.2% and the variation between triplicate samples of <1% (*10*), variations can be attributed to "real" variations and can be easily separated from the uncertainty due to chemical analysis.

A sampling frequency of 2 min was chosen in order to not miss any toilet flushes, and to cover a sufficiently long time period (4 h) with an acceptable total number of samples that require analysis ($n = 120$). This was based upon the following system evaluation. At the house, flushing a toilet lasts about 5−10 s. Flowing through the house connection this wastewater pulse expands to approximately 30 s to 5 min when entering the main sewer (*3*). Based on this initial duration, the sewer layout, and accounting for hydraulic conditions in the main sewer (effect of dispersion, ref *13*), it was estimated that an individual toilet flush would pass the influent of STP A within 2.5−10 min. While it was impossible to anticipate the timing of an MRI treatment and if the patient will be staying in the catchment or not, the most promising time to sample was thought to be on a Wednesday between 10 am and 2 pm (e.g., see Figure 2 in ref *9*): if the treatment was to take place in the early morning, considering the elimination half-life, the time span until the patient goes to the toilet, and travel times in the sewer system (estimated to be 1−2 h), the earliest water packets containing substantial amounts of Gd_{ant} could be expected approximately 3 h after treatment. If the treatment happened later in the afternoon, it was expected to still capture some water packets containing Gd_{anth} from the treatment(s) of the day before.

Different Composite Sampling Modes. The selection of sampling modes and frequencies was based on ref *2* to cover some of the most commonly applied *discrete* sampling modes and frequencies in STP or sewer studies (see Table 1). The following description outlines the concept to understand the required experimental setup since sampling uncertainty cannot be studied without making chemical analytical measurements (*14*).

The behavior of people and consumption of PPCPs in a catchment are not constant: pollutant loads (and concentrations) in a sewer vary from location to location and from day to day ("real" variation). Assuming that no perfect sample can be obtained in a monitoring campaign, sampling uncertainty will add some variation to the real variation. Furthermore, the uncertainty due to chemical analysis—including transport, preservation, storage, preparation, and instrumental error—again leads to additional variation. Since chemical analysis is the last step in the data generation process, the associated uncertainty can be determined directly. If a sample is properly preserved and/or analyzed straight after collection, the concentration of a compound in this sample can be treated as a constant: with independent replicates (subsamples) the precision of the analytical method can be quantified and with standard addition or by spiking labeled reference compounds the trueness can be enhanced. In contrast, the daily pollutant loads and the short-term concentration patterns in a sewer cannot be assumed as constants, which complicates the quantification of "sampling precision". Furthermore, it seems almost impossible to spike sewers with labeled reference compounds to generate a realistic PPCP pattern with a known daily mass to assess "sampling trueness". Therefore, we applied a *continuous flow-proportional* sampling mode (*15*) as a reference. Conceptually it is the most accurate (true and precise) sampling mode when sampling for loads of dissolved compounds. Instead

TABLE 1. Summary of Sampling Modes Applied at Locations A and B Representing Some of the Most Commonly Applied Discrete Sampling Modes and Frequencies According to a Review of 87 Journal Articles (*2*) (The Detailed Description and Visualization for Each Sampling Mode Can Be Found in Ref *2*. The Continuous Flow-Proportional Mode Was Applied As a Reference (See Text for More Details).)

	ID	mode[device]	sampling frequency	sample volume: individual, discrete (mL)	sample volume: total, composite (L)	number of individual samples pooled for a 24-h composite sample
location A (n = 5 days)	A1	flow-prop.[b]	continuous	0.7[a]	~10.9−18.6[c]	∞
	A2	volume-prop.[d]	1 per 400 m^3	200	~7.6−13.4[c]	~38−67[c]
	A3	time-prop.[f]	1 per 1 h	400	9.6	24
location B (n = 4 days)	B1	flow-prop.[b]	continuous	1.2[a]	~15.5−17.6[e]	∞
	B2	time-prop.[f]	1 per 20 min	100	7.2	72
	B3	time-prop.[f]	1 per 4 h	400	2.4	6
	B4	1 grab sample[g]	NA	1000	1	1

[a] Per m^3 wastewater in the sewer. [b] Device used: Watson Marlow 520UN (*15*). [c] Depending on the daily flow (min. 15,500 m^3 d^{-1}, max. 26,600 m^3 d^{-1}). [d] Device used: Sigma 900 MAX. [e] Depending on the daily flow (min. 13,000 m^3 d^{-1}, max. 14,700 m^3 d^{-1}). [f] Device used: Sentinel LPG/240. [g] Device used: scoop.

of a limited number of discrete samples it provides a properly weighted continuum. In *statistical terms*, this flow-proportional side stream represents the whole "population", and is not considered a (discrete) "sample" of the wastewater stream. Due to the expected, unknown short-term (minutes) and interday variations of pollutant loads, all sampling devices must be operated *simultaneously* to compare the different sampling modes. The devices were operated over five consecutive days at STP A and four days at STP B.

The compounds were selected to cover a wide range of different therapeutic classes and consumption, and were analyzed as described in SI 2 (samples STP A) and in ref *15* (samples STP B). After chemical analysis, the concentrations of all samples were normalized to the concentration obtained with the reference mode of the same day to compensate for day to day variation. It is assumed that sampling errors and analytical errors are independent. If the total variation of normalized concentrations exceeds the variation caused by chemical analysis, the contribution of the sampling step can be calculated as follows (variance decomposition):

$$s_{\text{sampling}} = \sqrt{s_{\text{total}}^{\,2} - s_{\text{chem.anal.}}^{\,2}} \quad (1)$$

where s_{sampling} is the variation attributable to sampling caused by the discrete sampling modes, s_{total} is the total variation of concentrations normalized to the reference mode, and $s_{\text{chem.anal.}}$ is the variation that was determined for chemical analysis (see SI 2 and Table SI 3A and B).

Modeling Study. Flows at STP influents are dramatically different depending on the sewer catchment and design, and can exhibit completely different patterns (see SI 1 and more examples in ref *2*). Ratios of minimum to maximum flows range from 2.5 to 20 (open channel, gravity flows) or are even more extreme in the case where intermittently operated pumps lead from no flow to maximum flows at highest pumping rates. Only at high temporal resolution does the flow pattern reveal the relevant information (see SI 1). It can usually be obtained from online measurements at the STP of interest. However, for this study, a realistic flow pattern was *modeled*, because no STP could provide sufficiently long time series at the required temporal resolution. The example modeled in this section is similar to the catchment and diurnal flow pattern of STP A.

The model presented in ref *3* was applied to generate a realistic load pattern (mg s^{-1}) which was then divided by the flow pattern (L s^{-1}) to calculate the concentration pattern (mg L^{-1}). The flow and load patterns were modeled at a temporal resolution of one second for a period of ten years. The load and corresponding concentration patterns were generated for 10, 100, 1000, and 10,000 wastewater pulses containing the substance of interest (individual pollutant peaks, e.g., toilet flushes, per day).

To determine the sampling error for a 24-h composite sample, a 24-h period starting at time t was randomly selected from the modeled flow and concentration time series. The "true" average concentration was calculated by multiplying the flow and concentration at highest temporal resolution (i.e., 1 s) divided by the total wastewater volume for the selected 24-h period (see eq 2). In eqs 3−5 the calculation procedures to mimic the different sampling modes are presented for a sampling frequency of one sample per hour.

$$\bar{C}_{\text{true}} = \frac{\sum_i c_i \cdot F_i}{V}, \quad V = \sum_i F_i$$
$$i = t, t+1, t+2, t+3, \ldots, t+86400 \text{ seconds} \quad (2)$$

$$\bar{C}_{\text{time-prop.}(\Delta T=1\text{ hour})} = \frac{1}{n}\sum_i c_i, \quad n = 24$$
$$i = t, t+3600, t+7200, \ldots, t+86400 \text{ seconds} \quad (3)$$

$$\bar{C}_{\text{volume-prop.}(\Delta V=1\text{-hour-eq.})} = \frac{1}{n}\sum_i c_i, \; x = \frac{V}{n}, \; n = 24$$
$$i = \text{points in time determined by flow variation every } x \text{ cubic meters} \quad (4)$$

$$\bar{C}_{\text{flow-prop.}(\Delta T=1\text{ hour})} = \frac{\sum_i c_i \cdot v_i}{\sum_i v_i}, \; v_i \propto F_i, \; n = 24$$
$$i = t, t+3600, t+7200, \ldots, t+86400 \text{ seconds} \quad (5)$$

where i is the time step, t is the start time of the 24-h period, c is the instantaneous concentration in a discrete sample, F is the flow in the sewer, $\bar{C}$ is the average concentration in the composite sample, and v is the volume of a flow-proportional discrete sample. The relative sampling error is calculated according to eq 6.

$$\text{relative sampling error} = \frac{\bar{C}_{\text{true}} - \bar{C}_{\text{sample}}}{\bar{C}_{\text{true}}} \quad (6)$$

Repeating this procedure for 1000 randomly selected, not overlapping 24-h periods was found to result in a stable

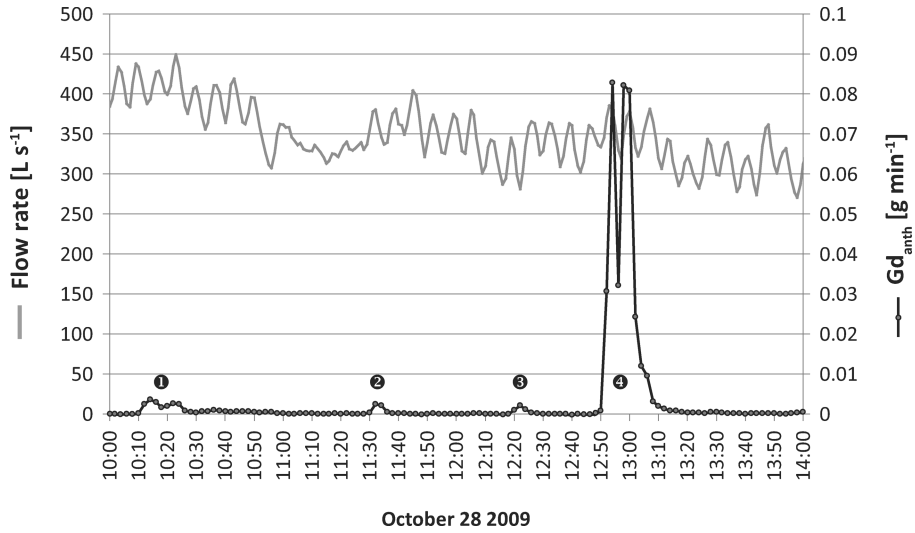

FIGURE 1. Flow rate (gray, measured online) and anthropogenic gadolinium (Gd_{anth}) loads (circles, grab samples at 2-min time intervals) in the influent of STP A. The flow slowly decreases from over 400 to 300 L s^{-1} over the sampling period. Short-term variations of the flow rate (approximately ±10%) are caused by the intermittent operation of the fine screen in the effluent of the primary clarifier where flow is measured. Therefore, these oscillations do not occur in the influent where samples were collected. The time shift of the flow signal was estimated to be less than 5 min and did not affect the sampling in the sewer.

sampling error distribution for each combination of sampling modes and sampling frequencies.

Results and Discussion

High-Frequency Grab Sampling. The pattern of anthropogenic gadolinium (Gd_{anth})—determined from 120 grab samples collected at 2-min time intervals in the influent of STP A and analyzed by ICP-MS (*10, 12*)—shows four distinct concentration peaks. The remaining samples have smooth rare earth element patterns with Gd_{anth} concentrations below the detection limit (loads are therefore $<4 \times 10^{-9}$ g min^{-1} (Figure 1)). The shapes and durations of peak 2 and 3 were as expected; they lasted 4−8 min. This is in the predicted range of 2.5−10 min. Peaks 1 and 4 extend over 15−20 min and reveal a double-peaked shape, each peak with similar maximum concentrations. It seems very unlikely that two independent toilet flushes, related to two different inhabitants in the catchment would temporally coincide and occur at almost the same time with the same load in the influent of the STP. A more plausible explanation was observed 500 m upstream from the STP where all sewage flows into a small chamber (a short pressurized section), entering at the bottom, flowing up to the surface before falling over a weir to the last section of pipe to the STP (again open channel gravity flow). Visual observations of large particulate matter confirmed the following, hydraulically variable behavior: when the water stream surfaces, some particles flow almost directly over the weir, while other particles are diverted to the rear part of the small retention basin. The wastewater from a toilet flush may therefore fall over the weir as a whole water packet (peaks 2 and 3), or may be split and partially retained before flowing farther downstream (peaks 1 and 4).

A typical dose of 15−30 mg kg^{-1} is standard for almost all contrast agents resulting in a total of 1−3 g Gd_{anth} applied to a 70−100 kg adult person. The total Gd_{anth} mass in peak 4 is approximately 0.7 g; relating to about $^1/_4$−$^1/_2$ of the mass typically administered to an adult patient; the corresponding treatment is expected to have taken place on the morning of the sampling day. The first three small peaks (p1 = 0.04 g, p2 = 0.01 g, p3 = 0.004 g) earlier in the day are likely to relate to the release of residual Gd_{anth} from the patient(s) treated on the previous day.

Most pharmaceutical residues (parent compounds and metabolites) are flushed into the sewer system via toilets, bathing water, and wastewater from washing machines (*16*). If only a few such wastewater pulses are expected over the course of a day, it is evident that a very high sampling frequency is necessary to capture these pulses; either to obtain a time series revealing the real short-term fluctuations, or to capture a representative composite sample from which it is possible to determine an average concentration and calculate a representative daily load; and even hundreds to thousands of pulses per day result in highly fluctuating patterns (see measured benzotriazole time series in ref *4*). Therefore, the main message is: if the dynamics for the substances of interest in the sewer system were not properly assessed in a pre-experiment, a precautionary high sampling frequency (<5 min) is necessary for minimizing sampling uncertainty. This also holds true for *composite* samples and more commonly used compounds (see next paragraph and refs *2 and 5* for more details and other solutions).

Different Composite Sampling Modes. In Figure 2, the results for three different compounds sampled for at the influent of STP B are plotted. The results for all compounds at STP B and STP A are summarized in the Table SI 3A and B. Subsequently all steps of the statistical evaluation are described for ranitidine (a histamine blocker) as an example: observed ranitidine concentrations were between 0.5 and 1.1 μg L^{-1} (Figure 2, left column). Observed day to day variations expressed as single standard deviations were between 25% and 28% for sampling modes B1, B3, and B4 and 10% for sampling mode B2 (time-proportional, $\Delta t = 20$ min). With sampling mode B1 (flow-proportional, continuous) the highest ranitidine concentration was measured on day 4, followed by the concentrations on day 1, 3, and 2. The reverse order of concentrations was obtained with sampling mode B4 (one grab sample per day).

Imagine the following situation: the four sampling devices were operated by four independent researchers, all analyzing their samples with the same method. Each researcher would have reached different conclusions for their studies, ranging from "no significant day to day variation" (B2), "highest concentration on day 4" (B1) to "highest concentration on day 2" (B4) or somewhere in between (B3). If a reader only read one study or did not know that the data were obtained at the same location over the same period—and consequently should have resulted in the same values on a particular day—probably all individual conclusions would have ap-

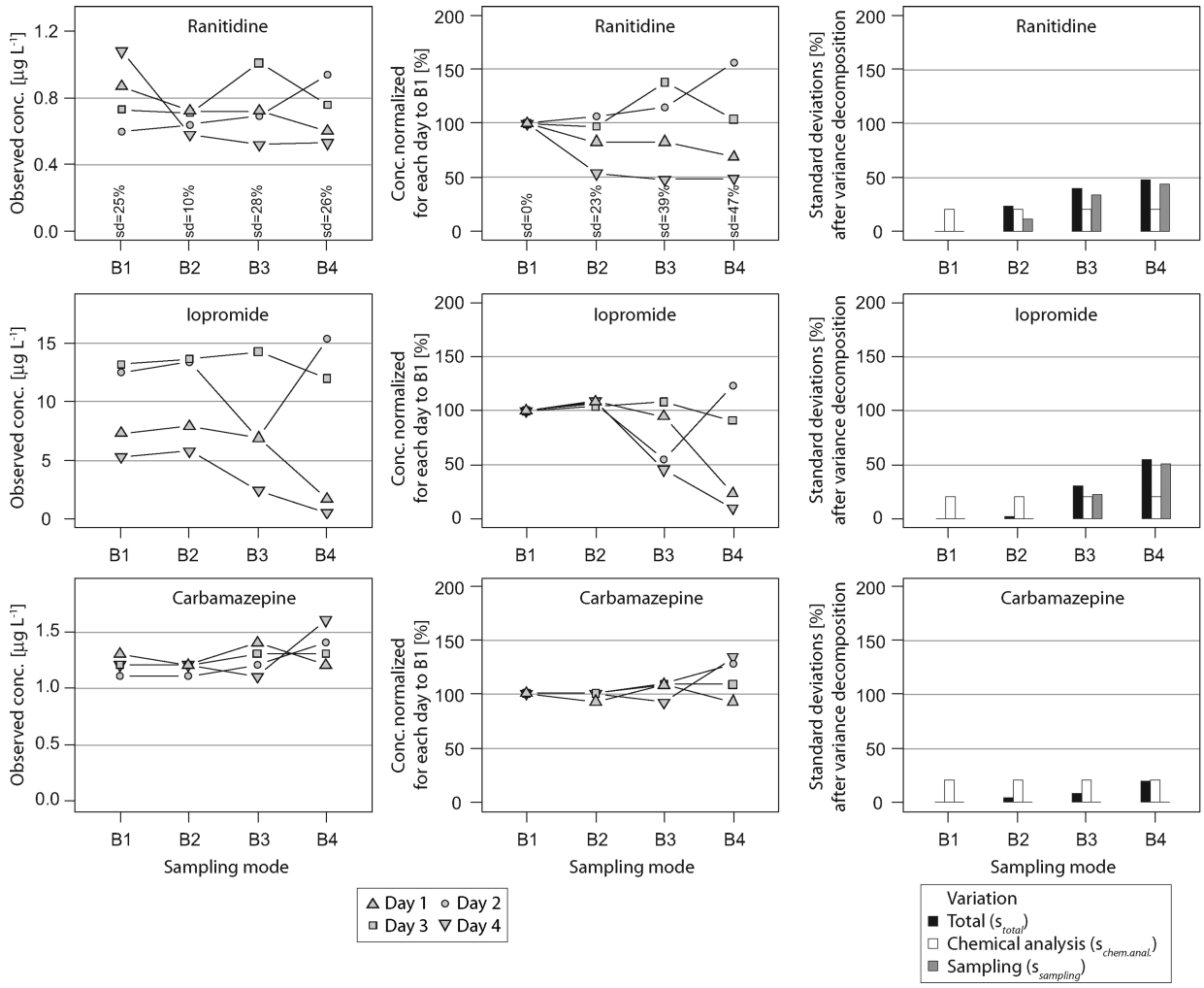

FIGURE 2. *x*-axis: Sampling modes: B1 = flow-proportional, continuous; B2 = time-proportional, Δt = 20 min; B3 = time-proportional, Δt = 4 h; B4 = one grab sample. Left: Observed concentrations for three compounds in the influent of sewage treatment plant B. Middle: Concentrations normalized for each day to sample B1. Right: Total variation (black), variation to be expected due to chemical analysis (white), and variation attributable to sampling artifacts (gray).

peared to be plausible. Unfortunately, most of the available current literature can be considered to "appear plausible" but there is insufficient evidence presented in most papers to evaluate whether the authors have actually captured a representative sample (*2*).

Conceptually, sampling mode B1 is the most appropriate to assess loads, as can be seen from eqs 2 and 5. Therefore, concentrations of each day were normalized to the concentration in the corresponding sample of B1 (see Figure 2, middle column). As per definition, the normalized B1 values are now all 100% and the corresponding concentration variation over the four days becomes zero. The variation of the normalized concentrations for the other sampling modes clearly increases with decreasing sampling frequency: 23% (B2), 39% (B3), and 47% (B4). The estimated variation caused by chemical analysis for ranitidine is 20% (see SI 2 and Table SI 3B). While the variation due to chemical analysis was "eliminated" with the normalization step for the B1 concentrations, it needs to be subtracted from the total variation according to eq 1 (variance decomposition). The remaining variation can now be attributed to sampling: 12% (B2), 34% (B3), and 43% (B4) as charted in Figure 2 (right column). The normalization step could also have been carried out with B2, B3, or B4. The results would have been in a similar range, revealing discrepancies among the different sampling modes (data not shown).

The evaluation of iopromide (an X-ray contrast media) shows similarities to and differences from the ranitidine example. The decreasing sampling frequency also results in a clear increase of variation. Furthermore, there is also one day (day 3) on which all sampling modes resulted in more or less the same concentrations. This shows that even a "lucky grab sample" or a low sampling frequency could have resulted in the same "average" concentration and the conclusion on day 3 would have been the same for all sampling modes. However, if sampling was only done on one day without the comparative sampling modes, we would not know if we were lucky or not. It is known that iopromide is expected to be excreted only by a very small number of patients in catchment B, and hence only a few wastewater pulses, i.e., toilet flushes, contain this substance (see also gadolinium as an MRI agent in the High-Frequency Grab Sampling section). Therefore, the systematic underestimation of iopromide concentrations is plausible: depending on the distribution of only a few peaks and the (small) number of samples pooled over a day, it is more likely to miss some peaks and sample more "zero concentrations" rather than hitting a peak concentration which could lead to an overestimation of the daily average concentration. In such cases even averaging long-term observations would not result in an unbiased estimate of real iopromide loads.

TABLE 2. Classification of Variations Caused by Sampling ($s_{sampling} = s_s$) for STP A and STP B (See Text for More Details)

	$s_s < 10\%$	$10\% < s_s \leq 20\%$	$20\% < s_s \leq 30\%$	$s_s > 30\%$
STP A total 60 cases (100%)	9 (15%)	20 (33%)	18 (30%)	13 (22%)
STP B total 69 cases (100%)	42 (61%)	9 (13%)	8 (12%)	10 (14%)

In the third example the total observed variation for carbamazepine (an antiepileptic drug) did not exceed the variation due to chemical analysis and hence, no variation could be attributed to sampling. Still, a consistent, slight increase of variation with decreasing sampling frequency can be observed. Although carbamazepine is typically widely applied—and therefore expected to be present in a large number of toilet flushes—it does not imply that the choice of sampling mode for carbamazepine does not need attention: (i) In a much smaller catchment there will only be a few pulses and sampling artifacts could occur (see also Modeling Study). (ii) The results at the influent of STP A show that even in a larger catchment a substantial part of the observed carbamazepine variation can be attributed to sampling uncertainty (Table SI 3A). The following two reasons are plausible explanations: (a) In the catchment and inlet works of STP A concentration variations are not attenuated due to retention basins or pumps; which contrasts with STP B, where $^1/_3$ of the wastewater is stored in a large buffer tank and all wastewater was pumped at least once (see SI 1). (b) Due to a change in the chemical analysis, the variation attributable to chemical analysis for carbamazepine at STP A was lower than for the samples from STP B (see SI 2). Therefore, almost all observed variation can be attributed to sampling uncertainty.

Thirty compounds were detected above the limit of quantification at STP A; for the two tested sampling modes A2 and A3 a total of 60 variations were calculated (deviation from the reference sampling mode A1). The variation caused by sampling was larger than 20% (up to a maximum of over 100%) in more than half of all cases (31) and only in 9 cases was it smaller than 10% (see Table 2). At STP B 23 compounds and three sampling modes resulted in 69 calculated variations; in $^1/_4$ of all cases (18) was the variation caused by sampling larger than 20%. Note: The presented values are only single standard deviations determined from samples of multiple days. Sampling errors for an individual day can be significantly larger (see also Figure 3)!

The individual values for the different compounds, locations, and sampling modes are listed in Tables SI 3A and B but are not so important since they cannot be directly transferred to any other sewer or influent of a STP. Sampling uncertainties for similar or other compounds highly depend on the site-specific boundary conditions. The main message is: different, commonly applied composite sampling modes which were presented as appropriate in previous literature to obtain a representative average sample from wastewater in a sewer are likely to lead to unreal variations (sampling artifacts), which is an unnecessary loss of environmental data quality.

Modeling Study. The error estimation model (*4*) mentioned in the Introduction suffers from two limitations: (i) The tested sampling intervals are shorter ($\Delta t = 1-30$ min) than the ones generally applied ($\Delta t \geq 30$ min) (*2*). (ii) The effects of flow variations were not assessed and only one sampling mode was tested (time-proportional). Therefore, longer sampling intervals and additional sampling modes were investigated in this study.

From the modeled ten-year period, 1000 independent 24-h sections were selected randomly. An example for a 24-h section of the flow pattern and two concentration patterns (approximately 10 and 10,000 individual pollutant peaks per day) are charted in Figure 3A. Relative sampling errors for the different sampling modes are charted in Figure 3B–D. Two general observations hold true for all sampling modes: (i) the sampling error increases with a decreasing number of wastewater pulses containing the substance of interest; and (ii) the sampling error also increases with decreasing sampling frequency. Note: The errors charted in Figure 3B–D are average values as single standard deviations from the error distributions. Therefore, the sampling error for one sample on an individual day can be significantly larger (see examples in Figure 3E–H).

The error distributions are only symmetric, and hence unbiased, for the flow- and volume-proportional mode (gray symbols in Figure 3C and D), except for the patterns with 10 pulses and long sampling intervals ($\Delta t \geq 1$ h). The time-proportional sampling mode does not weight individual subsamples according to the flow and is never unbiased for the presented combination of diurnal flow and concentration patterns (Figure 3A). As can be seen from the error distribution in Figure 3E pollutant loads are on average underestimated when flow and concentrations are positively correlated. This systematic error cannot be avoided with higher sampling frequencies in the time-proportional mode. The volume-proportional mode—commonly mistakenly referred to as flow-proportional (*2*)—does not weight the individual subsamples according to the flow either, but partially compensates for this deficiency with more samples during high flow, and less samples during low flow, respectively (compare Figure 3B–D at 10,000 pulses per day). Theoretically, the flow-proportional mode is the only mode properly weighting the individual subsamples of a composite sample. Its accuracy only depends on the sampling frequency. When sampling for a compound that is contained in only a small number of wastewater pulses, a discrete flow-proportional sampling mode with low sampling frequency on average also underestimates the load (white symbols in Figure 3D). For such compounds a very high sampling frequency is indispensable irrespective of flow pattern and sampling mode.

Gadolinium serves as a good example, since it was not only measured in the high-frequency sampling campaign but also for the comparison of different composite sampling modes at STP A (see Table SI 3A). The number of wastewater pulses per day containing Gd_{anth} is known to be between five and ten. The daily comparison of the volume-proportional mode (A2) with the continuous flow-proportional mode (A1) reveals the following experimentally derived differences due to sampling: −74%, +6%, +235%, +30%, and +36%. These values (stars in Figure 3G) resemble plausible realizations of the sampling error distribution, which is the most similar example from the modeling study. While the experimentally determined standard deviation for these five values is 114% (Table SI 3A) the average deviation expected from the model is 55% (black square in Figure 3C). The modeling study showed that approximately 1000 24-h periods need to be evaluated to derive a stable sampling error distribution. Therefore, it is obvious that four or five sampling days cannot describe the whole distribution. This is emphasized with the results from the time-proportional sampling mode. The experimentally derived differences compared to the reference sampling mode A1 for gadolinium are −55%, −40%, −45%, −25%, and −4% (stars in Figure 3E). Although the standard deviation of these values (20%) is five times smaller than expected (almost 100%, see circle in Figure 3B for 10 p d^{-1}) and suggests the experiment to be better than predicted with

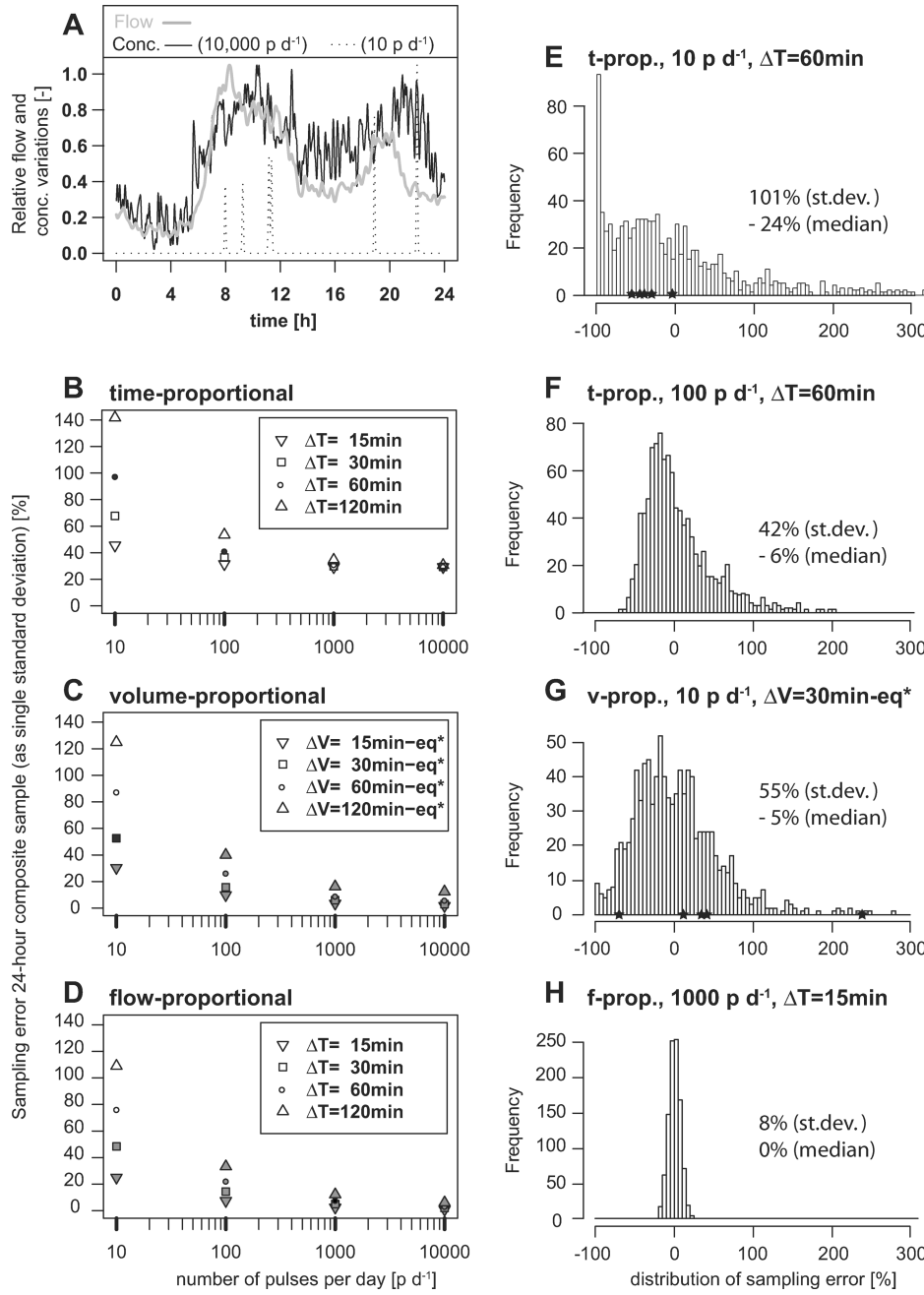

FIGURE 3. A: Flow and concentration patterns (approximately 10 and 10,000 pulses per day (p d^{-1})) for a 24-h section out of a ten-year period modeled at 1-s time intervals. B−D: Evaluation of sampling error for 1000 randomly selected 24-h periods for three different sampling modes and four different sampling frequencies. (*Note for C: In the case of the volume proportional sampling mode, the sampling frequency is variable; samples are taken every *x* cubic meters, during high flow more samples are collected than during low flow. To make a meaningful comparison *x* was selected to result in the same number of individual discrete samples per day as were used for the other two sampling modes with fixed time intervals.) Gray symbols indicate the combinations which lead to a symmetric, unbiased distribution (median = 0). E−H: Distribution of sampling errors for the combinations indicated with the solid black symbols in B−D (stars in E and G represent the experimentally determined sampling uncertainties for gadolinium at STP A, see text for more details).

the model, it is clear that the daily gadolinium loads were systematically underestimated (average −34%). While gadolinium may be seen as a "worst case" we selected it deliberately as an illustration through all approaches.

Transferability of Results to Other Catchments, Compounds, and Analytical Methods. The first two approaches presented in this study are not meant to be repeated in future studies due to the substantial analytical costs for high-frequency sampling and the high logistic effort for comparing different sampling modes. They were selected to demonstrate (i) the potential high (short-term) variability of PPCPs at the influent of a STP and (ii) the magnitude of sampling artifacts (from "not significant" to ">100%") when composite samples are obtained with traditional sampling modes and frequencies. In future studies sampling uncertainty should not be determined with a large effort, but minimized with a small effort. The required sampling frequency to maximize representativeness of composite samples is mainly influenced by the number of distinct wastewater packets in a sewer or at the influent of a STP. In open channel gravity sewers these water packets originate from household appliances (e.g., toilet flushes) and the number containing the substance of interest

depends on the number of users/patients. If this information can be collected for all compounds before sampling, and if the smallest number of pulses to be expected is known, the modeling approach can be used to determine the appropriate sampling frequency, ideally in a flow-proportional mode to properly account for flow variations (Figure 3D). If this information is not available before sampling, a precautionary high sampling frequency (<15 min) is advisible. In pressurized sewers the number of pulses is influenced by the (intermittent) operation of pumps and the frequency needs to be shorter than the shortest pump events that can be expected, in order to capture all water packets from whole subcatchments. A more detailed guide on selecting the appropriate sampling mode, frequency and other aspects related to sampling can be found in ref *2*.

When sampling for moving streams the whole cross section should be considered (*17*). However, no technical solution exists to perform this type of sampling, and we have to assume complete mixing of dissolved compounds; otherwise all PPCP sewer studies need to be questioned since sampling devices pump samples through a hose from one point in the cross section. For this study we focused on the analysis of the dissolved phase (assuming complete mixing over the whole cross section). To have a representative experimental reference, one set of samples was obtained with the *continuous* flow-proportional sampling mode. However, this mode may be less suitable for long-term monitoring because of the requirement for frequent sampler maintenance to minimize potential biofilm growth in the sampling hose. Furthermore, when sampling for particulate matter this mode cannot be applied meaningfully at all. Additional factors such as hose diameter and length, sampling velocity, distribution of particulate matter over the cross section (e.g., *18*), fractions to be analyzed for, dry vs wet weather flow, etc., and homogenization of sample before analysis must be considered (e.g., *19*). The resulting overall sampling and analytical uncertainty may be significantly larger than presented in this study. Furthermore, our experimental setup (simultaneous sampling at one location), chemical analysis (all samples analyzed in one run), and the comparison of ratios (relative differences) provided a unique approach to determine sampling uncertainty. How significant the sampling uncertainty will be in other sewer systems will highly depend on the catchment characteristics and a combination of sampling mode, frequency, and the method-specific analytical uncertainty for the compounds under investigation. This emphasizes the need for a thorough site- and case-specific systems analysis and is the responsibility of each research team. Sampling must be an informed choice and should not only be based on the convenience of an available device. The examples presented in this paper demonstrate that sampling uncertainty exists, and can be relevant in the context of evaluating full-scale PPCP fluxes in sewers. The effort to reduce sampling uncertainty is relatively small, and should be undertaken irrespective of how it compares to analytical uncertainty. This will improve the reproducibility of results and avoid (potentially unrecognizable) sampling artifacts.

Acknowledgments

We thank the Regional Councils Moreton Bay and Toowoomba and the personnel operating the South Caboolture STP and Wetalla STP for their interest in our research and their logistical and practical assistance. Furthermore, we acknowledge the financial support by the Urban Water Security Research Alliance and the Swiss National Science Foundation (Grant PBEZP2-122958 awarded to C. Ort).

Supporting Information Available

Details on the two catchments under investigation, the method for chemical analyses, and a table with the results for more compounds from samples obtained from STP A and B for the section Different Composite Sampling Modes. This information is available free of charge via the Internet at http://pubs.acs.org.

Literature Cited

(1) *ISO Water quality - Sampling - Part 1: Guidance on the design of sampling programmes*; ISO 5667-1; ISO: Genève, Switzerland, 1980.
(2) Ort, C.; Lawrence, M. G.; Rieckermann, J.; Joss, A. Sampling for PPCPs and illicit drugs in wastewater systems: Are your conclusions valid? A critical review. Environ. Sci. Technol. in press, doi: 10.1021/es100779n.
(3) Ort, C.; Schaffner, C.; Giger, W.; Gujer, W. Modeling stochastic load variations in sewer systems. *Water Sci. Technol.* **2005**, *52*, 113–122.
(4) Ort, C.; Gujer, W. Sampling for representative micropollutant loads in sewer systems. *Water Sci. Technol.* **2006**, *54*, 169–176.
(5) Clesceri, L. S.; Greenberg, A. E.; Eaton, A. D. *Standard Methods for the Examination of Water and Wastewater*; American Public Health Association: Washington, DC, 1998; ISBN 0875532357, Vol. 20.
(6) *ISO Water quality - Sampling - Part 10: Guidance on sampling of waste waters*; ISO 5667-10; ISO: Genève, Switzerland, 1992.
(7) *U.S. EPA Handbook for Sampling and Sample Preservation of Water and Wastewater*; 600/4-82-029; Environmental Monitoring and Support Laboratory, 1982.
(8) Miège, C.; Choubert, J. M.; Ribeiro, L.; Eusèbe, M.; Coquery, M. Fate of pharmaceuticals and personal care products in wastewater treatment plants - Conception of a database and first results. *Environ. Pollut.* **2009**, *157*, 1721–1726.
(9) Bau, M.; Dulski, P. Anthropogenic origin of positive gadolinium anomalies in river waters. *Earth Planet. Sci. Lett.* **1996**, *143*, 245–255.
(10) Lawrence, M. G.; Bariel, D. G. Tracing treated wastewater in an inland catchment using anthropogenic gadolinium. Chemosphere, in press, doi: 10.1016/j.chemosphere.2010.05.001.
(11) Rauch, W.; Brockmann, D.; Peters, I.; Larsen, T. A.; Gujer, W. Combining urine separation with waste design: an analysis using a stochastic model for urine production. *Water Res.* **2003**, *37*, 681–689.
(12) Lawrence, M. G.; Greig, A.; Collerson, K. D.; Kamber, B. S. Direct quantification of rare earth element concentrations in natural waters by ICP-MS. *Appl. Geochem.* **2006**, *21*, 839–848.
(13) Rieckermann, J.; Neumann, M.; Ort, C.; Huisman, J. L.; Gujer, W. Dispersion coefficients of sewers from tracer experiments. *Water Sci. Technol.* **2005**, *52*, 123–133.
(14) Ramsey, M. H.; Thompson, M. Uncertainty from sampling, in the context of fitness for purpose. *Accredit. Qual. Assur.* **2007**, *12*, 503–513.
(15) Ort, C.; Lawrence, M. G.; Reungoat, J.; Eaglesham, G.; Carter, S.; Keller, J. Determining the fraction of pharmaceutical residues in wastewater originating from a hospital. *Water Res.* **2010**, *44*, 605–615.
(16) Daughton, C. G.; Ruhoy, I. S. Environmental footprint of pharmaceuticals: the significance of factors beyond direct excretion to sewers. *Environ. Toxicol. Chem.* **2009**, *28*, 2495–2521.
(17) Gy, P. *Sampling for Analytical Purposes*; John Wiley & Sons Ltd.: New York, 1998; ISBN 0 471 97956 2.
(18) Larrarte, F. Suspended solids within sewers: an experimental study. *Environ. Fluid Mech.* **2008**, *8*, 249–261.
(19) Clark, S. E.; Siu, C. Y. S. Measuring solids concentration in stormwater runoff: Comparison of analytical methods. *Environ. Sci. Technol.* **2008**, *42*, 511–516.

ES100778D