UNCERTAINTY IN MEASURED SEDIMENT AND NUTRIENT FLUX IN RUNOFF FROM SMALL AGRICULTURAL WATERSHEDS

R. D. Harmel, K. W. King

ABSTRACT. *Storm water quality sampling techniques vary considerably in the resources required for sample collection and analysis, and potentially in the resulting constituent flux estimates. However, quantitative information on sampling error is rarely available for use in selecting appropriate sampling techniques and for evaluating the effects of various techniques on measured results. In an effort to quantify uncertainty in constituent flux measurement for flow-interval sampling techniques, water quality data were collected from two small watersheds in central Texas. Each watershed was instrumented with two automated samplers. One sampler was programmed to take high-frequency composite samples to determine the actual load for each runoff event. The other sampler collected discrete samples, from which 15 strategies with 1.32 to 5.28 mm volumetric depth sampling intervals with discrete and composite sampling were produced. Absolute errors were consistently larger for suspended sediment than for NO_3-N and PO_4-P for both individual event and cumulative loads, which is attributed to differences in the variability of within-event constituent concentrations. The mean event-specific coefficient of variation (CV) ranged from 0.53 to 0.69 for sediment, from 0.38 to 0.39 for NO_3-N, and from 0.18 to 0.21 for PO_4-P. Event-specific CV values were correlated with the magnitude of absolute errors for individual event loads, with mean r values of 0.52 and 0.57 for the two sites. Cumulative errors were less than ±10% for all sampling strategies evaluated. Significant differences in load estimate error resulted from changes in sampling interval, but increasing the number of composited samples had no effect; therefore, composite sampling is recommended if necessary to manage the number of samples collected.*

Keywords. *Nitrogen, Nonpoint-source pollution, Phosphorus, Sampling error, Water quality sampling.*

Nonpoint-source (NPS) pollution contributes to water quality impairment in many U.S. water bodies (USEPA, 1995, 2000). Thus, researchers and regulatory personnel are forced to monitor storm runoff to quantify, understand, and mitigate detrimental impacts of NPS pollution on water quality. Recent research has focused on aspects of storm sampling such as strategy selection, sampler operation, and error comparison (Claridge, 1975; Richards and Holloway, 1987; Wells et al., 1990; Rekolainen et al., 1991; Izuno et al., 1998; Miller et al., 2000; Stone et al., 2000; McFarland and Hauck, 2001; Guo et al., 2002; Harmel et al., 2002; Argourdis and Edwards, 2003; Harmel et al., 2003b; King and Harmel, 2003). In their 2003 study, Agouridis and Edwards emphasized that the collection and analysis of storm samples is a difficult, time consuming, and expensive task. This simple statement emphasizes the problematic situation faced in sampling projects, which is to balance adequate characterization of water quality (frequent sample collection throughout the event duration) with limited sample analysis resources and with equipment constraints (sampling capacity). Although uncertainty impacts the results of water resource applications and, therefore, should be an important consideration, it is often overlooked because little is known about the effects of sample collection techniques on sampling error.

One aspect of water resource management desperately needing further research is the analysis of uncertainty in constituent transport measurements, which applies most directly to water quality modeling. In Total Maximum Daily Load (TMDL) projects, for example, water quality monitoring data are often used to calibrate watershed models, which are then used to estimate loads from various sources. TMDLs include a margin of safety to account for uncertainty in estimations, but research by Hession et al. (1996) and Haggard et al. (2003) indicates the need for uncertainty or risk analysis in TMDL determinations. An important component of uncertainty analysis is error propagation, which begins with fundamental data; in this case, uncertainty in load values calculated from automated water quality sampling.

Most automated sampling strategies use time or flow intervals to determine when samples are taken. With time-interval strategies, samples are typically taken on uniform time intervals (for example every 30 min) or variable time intervals (typically with more frequent samples initially, and then less frequent as the storm proceeds). With flow-interval strategies, samples are taken on uniform flow intervals (such as every 2000 m^3 or 2.5 mm volumetric depth). Statistical sampling theory indicates that smaller sampling intervals (the more samples taken) produce better estimates of the actual population characteristics (Haan, 2002). This theory is supported in storm monitoring by a

Article was submitted for review in June 2004; approved for publication by the Soil & Water Division of ASABE in July 2005.

Mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA.

The authors are **R. Daren Harmel, ASABE Member Engineer,** Agricultural Engineer, USDA-ARS, Temple, Texas; and **Kevin W. King, ASABE Member Engineer,** Agricultural Engineer, USDA-ARS, Columbus, Ohio. **Corresponding author:** R. Daren Harmel, USDA-ARS, 808 E. Blackland Rd., Temple, TX 76502; phone: 254-770-6521; fax: 254-770-6561; e-mail: dharmel@spa.ars.usda.gov.

limited number of analytical studies, including Richards and Holloway (1987), Shih et al. (1994), Miller et al. (2000), and King and Harmel (2003), and by a small plot runoff study (King and Harmel, 2004). However, to our knowledge, no field studies have attempted to quantify sampling error in measured storm water quality data. It is important to note that sampling error is defined as sampling variability or sampling uncertainty and does not include mistakes in data collection and processing (Haan, 2002). Based on this need for field evaluations of sampling error, this study was designed to evaluate uncertainty in load estimates for several flow-interval (also referred to as flow-weighted, flow-stratified, or flow-proportional) sampling strategies. Specifically, the objective of this study was to compare actual suspended sediment and dissolved NO_3-N and PO_4-P loads measured with an intensive sampling strategy to loads estimated with various discrete and composite flow-interval strategies.

METHODS

SITE DESCRIPTION

Since the late 1930s, hydrologic data have been collected at the USDA-ARS Grassland Soil and Water Research Laboratory near Riesel, Texas, which make it one of the longest continuously active, intensively monitored hydrological research sites in the U.S. Soils at the site are dominated by Houston Black clays, which are classic Vertisols noted for their strong shrink/swell potential. Land surface slopes at the site range from 2% to 4%.

Two watersheds at the Riesel facility, Y13 and Y, were utilized in this study. The outlet of watershed Y13 is an “edge of field” station that receives runoff from a 4.6 ha cultivated field. The outlet of Y receives runoff from a 125.1 ha mixed land use watershed, which includes Y13. A flow control structure with a well-established, reliable stage discharge relationship is located at each watershed outlet (Y13 - 5:1 broad-crested v-notch weir; Y - Columbus A-1 deep-notch weir). The flow control structures provide reliable flow data for accurate flow-interval sampling.

DATA COLLECTION

Two years of data were collected for this study (2003-2004). Each flow control structure was equipped with two ISCO 6700 samplers (ISCO, Inc., Lincoln, Neb.). One sampler with a single sample collection bottle (16 L) was programmed to collect composite 200 mL samples taken at 1.32 mm flow intervals throughout runoff events (defined as flow greater than 0.14 m^3/s for Y and greater than 0.02 m^3/s for Y13). Referring to discharge intervals in volumetric depth units (mm), which represent mean runoff depth over the entire watershed, as opposed to volume units (m^3) normalizes discharge over various watershed sizes. This notation allows a consistent transfer of methods and results to watersheds of differing size. In this study, 1.32 mm flow intervals are equal to 1650.8 m^3 for Y and 60.4 m^3 for Y13. This intensive sampling strategy was assumed to quantify the actual load for each runoff event. With this strategy, large runoff events (up to 106 mm) were completely sampled. The concentration from the single composite sample represented the event mean concentration (EMC), which is the arithmetic mean of sample concentrations collected on equal discharge intervals. The load was then determined by multiplying the EMC and the runoff volume.

The second sampler at each site was equipped with a 24-bottle (1 L) arrangement and programmed to take discrete, 1.32 mm flow-interval samples. For each runoff event, these discrete samples were used to calculate loads for 15 sampling strategies, which included sampling intervals of 1.32, 2.64, and 5.28 mm along with discrete sampling and compositing with two to five samples per bottle. With this design, runoff events of less than 31.7 mm were completely sampled, but only the first 31.7 mm of runoff was sampled in larger-volume events. The samples used to calculate loads for the discrete and composite 1.32 mm flow-interval strategies are shown in table 1, and the same procedure was used to calculate loads for the 2.64 mm and 5.28 mm strategies.

JUSTIFICATION OF STUDY DESIGN AND ASSUMPTIONS

An important factor to consider when using automated sampling equipment is uniformity of water quality across the flow cross-section and within the water profile. It is generally assumed that water quality can be adequately sampled at a single intake point in small streams because of well-mixed conditions. This assumption is generally valid for dissolved constituents in small streams and in larger streams when not immediately downstream of significant point sources (Martin et al., 1992; Raymond Slade, personal communication, 2004). The validity of this assumption is much more limited in terms of sediment transport. Even in small streams dominated by fine particles, vertical and horizontal gradients can occur in sediment concentration. In large streams and rivers and in conditions of coarse particle transport, sediment concentrations vary widely vertically and horizontally; therefore, alternative techniques should be used to quantify sediment transport in these conditions. At this study site, it is valid to assume uniformity of dissolved and particulate

Table 1. Sample (bottle) numbers used to calculate estimated loads for the 1.32 mm flow-interval strategies.

Runoff (mm)	Discrete	No. of Composite Samples per Bottle			
		2	3	4	5
1.32	1				
2.64	2	1,2			
3.96	3		1,2,3		
5.28	4	3,4		1,2,3,4	
6.60	5				1,2,3,4,5
7.92	6	5,6	4,5,6		
9.24	7				
10.56	8	7,8		5,6,7,8	
11.88	9		7,8,9		
13.20	10	9,10			6,7,8,9,10
14.52	11				
15.84	12	11,12	10,11,12	9,10,11,12	
17.16	13				
18.48	14	13,14			
19.80	15		13,14,15		11,12,13,14,15
21.12	16	15,16		13,14,15,16	
22.44	17				
23.76	18	17,18	16,17,18		
25.08	19				
26.40	20	19,20		17,18,19,20	16,17,18,19,20
27.72	21		19,20,21		
29.04	22	21,22			
30.36	23				
31.68	24	23,24	22,23,24	21,22,23,24	

concentrations because the watersheds are clay-dominated with well-mixed conditions at the sampling point.

To quantify load estimate errors, actual loads must first be determined with an intensive strategy that samples on small intervals. With a sufficiently frequent sampling intensity, the assumption can be made that the measured load is equivalent to the actual load. Then, load estimates from various sampling strategies can be compared to the actual load. For flow-interval sampling, small flow intervals best represent storm loads (Richards and Holloway, 1987; Miller et al., 2000; King and Harmel, 2003). King and Harmel (2003) concluded that flow-interval samples should be taken on intervals less than 2.5 mm to capture actual loads. In the present study, we assume that samples taken on 1.32 mm sampling intervals are sufficient to calculate the actual load for each storm. This assumption is required because it is impractical in field studies to capture the entire runoff volume for actual load measurement, which has been known for some time to create difficulty in load determination (Parsons, 1954). The 1.32 mm sampling interval approaches the practical sampling limit because sample pumps require 1 to 2 min to purge the sample line, collect a sample, and then reclear the sample line. For the 300 storms analyzed by King and Harmel (2003), only 1% of the storms had peak flow rates greater than 1 mm/min volumetric depth (a range of 0.003 to 202.5 m^3/s).

Statistical Analysis

A variety of statistical methods were used to analyze constituent behavior and to evaluate the performance of the 15 sampling strategies in quantifying event and cumulative loads. The correlation coefficient (r) was used to evaluate the relationship between flow rate and constituent concentration, and the coefficient of variation (CV) was used to measure the dispersion of within-event concentrations (Haan, 2002). Absolute and relative errors were determined between actual and estimated individual event loads and cumulative loads by the following equations:

$$ABS = L_{ac} - L_{est} \tag{1}$$

$$REL = (L_{ac} - L_{est}) \times 100 / L_{ac} \tag{2}$$

where

ABS = absolute error (kg/ha)
REL = relative error (%)
L_{ac} = actual load (kg/ha)
L_{est} = estimated load (kg/ha).

Resulting negative values from these equations represent overestimation, and positive values represent underestimation. Potential differences between actual and estimated loads were evaluated with Mann-Whitney tests (medians) and paired t-tests (means). Analysis of variance (ANOVA) was used to determine the effects of sampling strategy on root mean square error (RMSE), error sum of squares (errorSS), and mean and maximum error for event loads and on cumulative error for accumulated loads. All statistical tests were performed with Minitab 13 software and procedures described in Minitab (2000), Helsel and Hirsch (1993), and Haan (2002). Tests of significance were conducted at an *a priori*, $\alpha = 0.05$ probability level.

Results and Discussion

Characteristics of Runoff Events

The 20 runoff events that occurred at Y and the 17 runoff events at Y13 in 2003 and 2004 were used to compare actual loads to loads estimated from various flow-interval sampling strategies. Differing hydrologic conditions produced relatively few runoff events in 2003 and frequent events in 2004. The greatest monthly June and November precipitation totals, since data collection at the Riesel site began in 1938, were measured in 2004. June received 358 mm of precipitation, and November received 263 mm. Runoff event rainfall totals ranged from 18 to 128 mm, and runoff depths ranged from 1 to 98 mm. The return period of the maximum daily rainfall recorded during the study (101 mm) is approximately three years (Harmel et al., 2003a).

Constituent Behavior in Relation to Flow

In general, sediment concentrations closely followed the trends in flow rate, whereas NO_3-N and PO_4-P concentrations decreased as flow increased (fig. 1). Representative events for each watershed are shown in figures 2a and 2b to illustrate the relationship of constituent concentrations to flow rate.

Many studies, including the present one and others such as Robertson and Roerish (1999) and Tate et al. (1999), have documented that the peak concentration of sediment and sediment-derived constituents typically coincides with or precedes peak flow and that these constituents exhibit a hysteresis effect during decreasing flow (concentrations at a particular flow rate on the falling limb of the hydrograph are typically lower than at the same flow rate on the rising limb). The positive correlation between flow and sediment concentrations (fig. 1, table 2) was weaker and more variable for watershed Y13 (mean r = 0.33) than for Y (mean r = 0.56), possibly due to the more pronounced influence of temporal changes in cover conditions for the cultivated watershed (Y13). Sediment concentrations also varied substantially within runoff events. The largest range of sediment concentrations within a runoff event was 350 to 3328 mg/L for Y13 and 384 to 1953 mg/L for Y. The CV values for sediment averaged 0.53 for Y13 and 0.69 for Y (table 2).

In contrast, NO_3-N concentrations tended to be negatively correlated with flow (fig. 1, table 2), especially for Y13 with a mean r value of −0.31. For Y, the correlation between NO_3-N and flow was weaker (mean r = −0.13), which is attributed to various mechanisms of NO_3-N transport (lateral

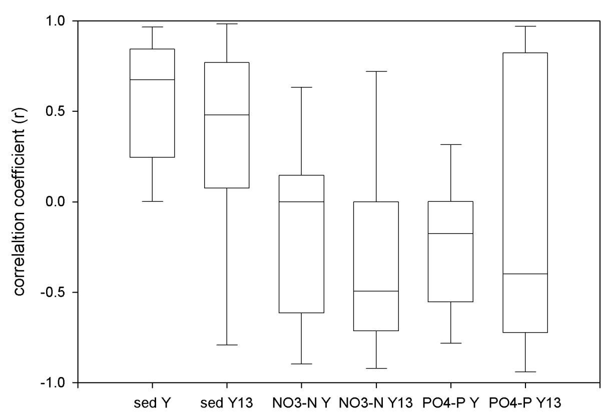

Figure 1. Distribution of correlation coefficients between flow (Q) and sediment, NO_3-N, and PO_4-P concentrations.

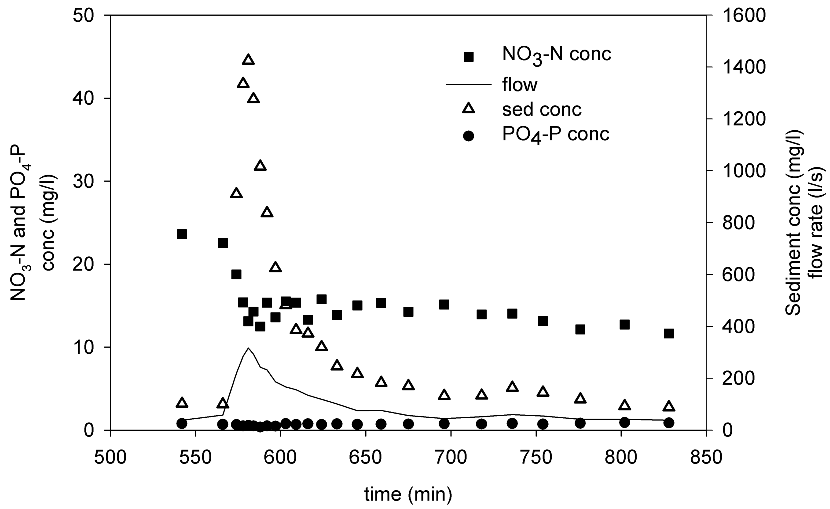

Figure 2a. Relationship between flow rate (Q) and sediment, NO_3-N, and PO_4-P concentrations for watershed Y13 on 9 October 2003.

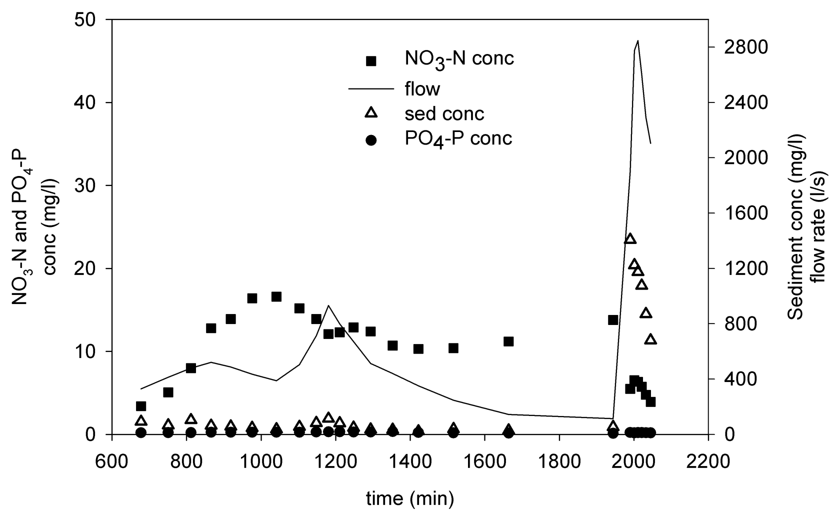

Figure 2b. Relationship between flow rate (Q) and sediment, NO_3-N, and PO_4-P concentrations for watershed Y on 20 February 2003.

subsurface return flow and surface runoff) and multiple sources of NO_3-N (fertilizer application and manure deposition). The largest range of NO_3-N concentrations within a runoff event was 28 to 88 mg/L for Y13 and 0.6 to 17 mg/L for Y, which is much lower than for sediment. The CV values for runoff events averaged 0.38 and 0.39 for Y13 and Y, respectively (table 2).

Table 2. Summary statistics of event-specific relationships between constituent concentration (mg/L) and flow rate (Q, m³/s) and constituent concentration variability.

	Correlation Coefficient (r)			Coefficient of Variation (CV)		
	Q/NO_3-N	Q/PO_4-P	Q/Sed.	NO_3-N	PO_4-P	Sed.
Watershed Y13						
Max.	0.80	0.98	0.98	0.98	0.37	1.15
Min.	−1.00	−1.00	−0.94	0.00	0.03	0.24
Mean	−0.31	−0.12	0.33	0.38	0.18	0.53
Median	−0.50	−0.40	0.48	0.34	0.17	0.44
Watershed Y						
Max.	0.75	0.36	0.99	1.25	0.43	1.49
Min.	−0.98	−0.88	−0.14	0.00	0.05	0.34
Mean	−0.13	−0.24	0.56	0.39	0.21	0.69
Median	0.00	−0.18	0.67	0.30	0.20	0.61

Similar to NO_3-N, PO_4-P concentrations tended to be negatively correlated with flow (fig. 1, table 2), with mean r values of −0.12 and −0.24 for Y13 and Y, respectively. Again, the relationship between flow and constituent concentration varied more at Y13 than at Y, possibly due to changes in soil nutrients available for transport as a function of time relative to fertilizer application. PO_4-P concentrations exhibited little variability within runoff events, as shown by the maximum ranges of 0.34 to 0.96 mg/L for Y13 and 0.49 to 0.88 mg/L for Y. Mean CV values for PO_4-P (0.18 and 0.21) were lower than for sediment or NO_3-N (table 2).

It is difficult to generalize sampling strategy performance for various runoff events because of differing behavior for individual constituents, and the substantial variability in runoff event duration, intensity, and flow patterns. The temporal variability between runoff and pollutant transport (due to changes in surface vegetation, time since fertilizer application, rainfall intensity, runoff/precipitation ratio, etc.) also adds to the difficulty, but in spite of this variability, two general patterns were evident. As expected, differences in sampling strategy performance were most evident in large-volume events, as only the first 31.7 mm of runoff was sampled by the 24-bottle 1.32 mm discrete strategy from which the other 14 strategies were represented. In these events, the over- or under−estimation of loads was typically driven by the timing of peak constituent concentrations in relation to the period when samples were collected. It was also evident that the effect of not sampling the peak transport period was greater for sediment loads than for PO_4-P loads because sediment concentrations varied more within runoff events.

Event Loads

The differences between actual and estimated sediment, NO_3-N, and PO_4-P loads for individual runoff events were not normally distributed based on the Kolmogorov-Smirnoff test (Haan, 2002). Therefore, nonparametric Mann-Whitney tests were used to analyze median loads for each site (Helsel and Hirsch, 1993). Paired t-tests were also used to compare mean loads, as if the common assumption of normality was valid. Results from these tests indicated no significant difference in mean (or median) sediment, NO_3-N, and PO_4-P loads between the actual load and estimated loads for the 15 sampling strategies evaluated. However, the accuracy for individual events or for accumulated events over a period of interest has greater practical importance in terms of sampling strategy performance than does the accuracy of mean (or median) loads.

Absolute errors for individual events were quite different for the three constituents. Errors in sediment load estimation for individual events ranged from −157.9 kg/ha (overestimation) to +76.4 kg/ha (underestimation) for Y13 and from −38.7 to +39.1 kg/ha for Y. The magnitudes of errors were smaller for NO_3-N loads at each site (−2.0 to +0.9 kg/ha for Y13, and −0.4 to +0.7 kg/ha for Y). Error magnitudes were smaller yet for PO_4-P loads (−0.06 to +0.04 kg/ha for Y13, and −0.04 to +0.01 kg/ha for Y).

Cumulative Loads

The cumulative effect of errors for individual events may be even more important, as constituent loads are often reported on total mass lost per year or project duration. Because absolute errors in individual events were larger for sediment loads than for NO_3-N and PO_4-P loads, absolute errors in cumulative loads followed the same pattern. Errors in cumulative sediment load estimation for specific strategies ranged from −7.2 to −231.5 kg/ha for Y13 and from −57.0 to +19.2 kg/ha for Y. Cumulative NO_3-N load errors ranged from −0.3 to −3.6 kg/ha and from −0.5 to +0.2 kg/ha for Y13 and Y, respectively. Errors in cumulative PO_4-P load estimation were −0.04 and +0.04 kg/ha for Y13 and −0.08 and −0.10 kg/ha for Y.

Although absolute errors in cumulative load estimation were largest for sediment, less for NO_3-N, and least for PO_4-P, the ranges of relative errors were quite similar. Relative errors in cumulative load estimation ranged from −9.1% to +2.7% for sediment (fig. 3), from −9.2% to +2.0% for NO_3-N (fig. 4), and from −9.2% to +2.0% for PO_4-P (fig. 5).

Effects of Within-Event Concentration Variability on Sampling Error

The difference in sampling strategy performance for the three constituents studied has important ramifications for storm water quality sampling. Absolute errors in individual event and cumulative load estimation were largest for sediment and least for PO_4-P. This ranking of load estimate errors (sediment > NO_3-N > PO_4-P) is attributed to differences in within-event concentration variability. The mean CV across sites for within-event concentrations was 0.61 for sediment, 0.39 for NO_3-N, and 0.19 for PO_4-P. Similarly, the largest within-event range of sediment concentrations was 2977 mg/L compared to a range of 60 mg/L for NO_3-N and only 0.62 mg/L for PO_4-P.

This observation that constituents with the largest within-event concentration variability will experience the largest errors in load estimation, which was also noted by Claridge (1975) and Rekolainen et al. (1991), was confirmed by comparing event-specific CV values to the magnitude of absolute errors for each sampling scheme. With data pooled for all three constituents on both watersheds, significant correlation existed between CV values and event-specific errors (mean r value = 0.47). When the sites were separated, the mean r value increased to 0.52 for Y13 and increased to 0.57 for Y. While this correlation does not establish a cause/effect relationship, it is consistent with accepted statistical sampling theory. For populations with a higher variance, more samples are necessary to estimate the

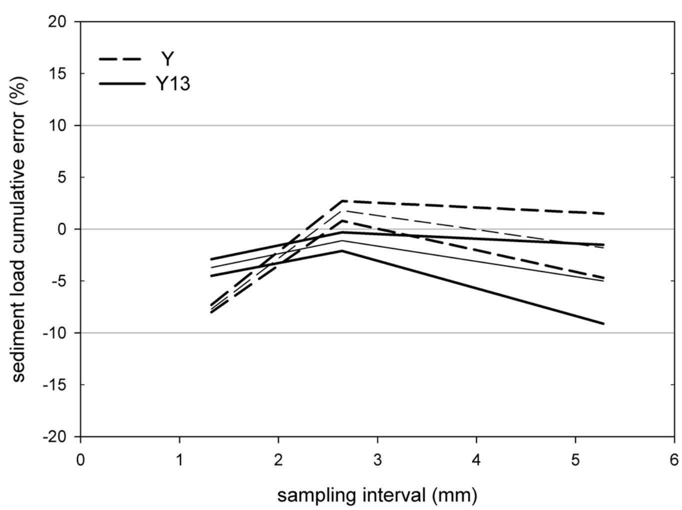

Figure 3. Cumulative error in sediment load determination as a funciton of sampling interval for discrete and composite sampling; the maximum, minimum, and average percent error are presented.

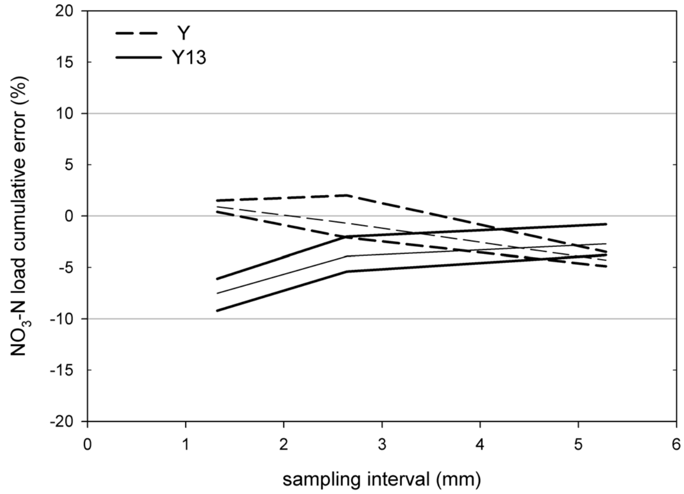

Figure 4. Cumulative error in NO_3-N load determination as a funciton of sampling interval for discrete and composite sampling; the maximum, minimum, and average percent error are presented.

population mean (in this case EMC) within a given probability of error (Haan, 2002). Whereas patterns in absolute errors were related to differences in within-event concentration variability, relative errors were not affected by differences in constituent behavior. Relative errors in cumulative loads were between −9.2% and +2.7% for all constituents (figs. 3 through 5).

Effects of Various Flow-Interval and Composite Sampling Options on Sampling Error

In the analysis of event-specific and cumulative loads, it became apparent that sampling interval (frequency) affected sampling error but that the number of composite samples had no effect. Selected indicators of error for each of the 15 sampling strategies evaluated demonstrate this difference (table 3). For watershed Y, ANOVA indicated that increasing the sampling interval resulted in significant differences in RMSE, errorSS, cumulative error, and mean and maximum event-specific error for all three constituents. For watershed Y13, significant differences were determined for all of these measures of error except for NO_3-N errorSS. Although these statistically significant differences were experienced for all constituents, it should be kept in mind that absolute differences were largest for sediment and least for PO_4-P. In contrast, ANOVA resulted in no significant differences in sampling error based on the number of composite samples.

Relevance to Sampling Strategy Development

The need for a balance between limited sampling and analysis resources and accurate characterization of storm water quality has been recently expressed (Shih et al., 1994; Agouridis and Edwards, 2003; Harmel et al. 2003b); however, the effects of sample collection techniques on measured results are rarely evaluated. To achieve this balance, it is typically necessary to manage the number of samples collected without increasing uncertainty. This involves setting the sampling interval and the number of composite samples to adequately capture constituent behavior (such as first flush) without exceeding sampler capacity prior to the end of large-volume events. Based on the finding that sampling error is affected by increasing the sampling interval but not by increasing the number of composited samples, the number of samples collected should be managed by adjusting the number of composite samples, not the

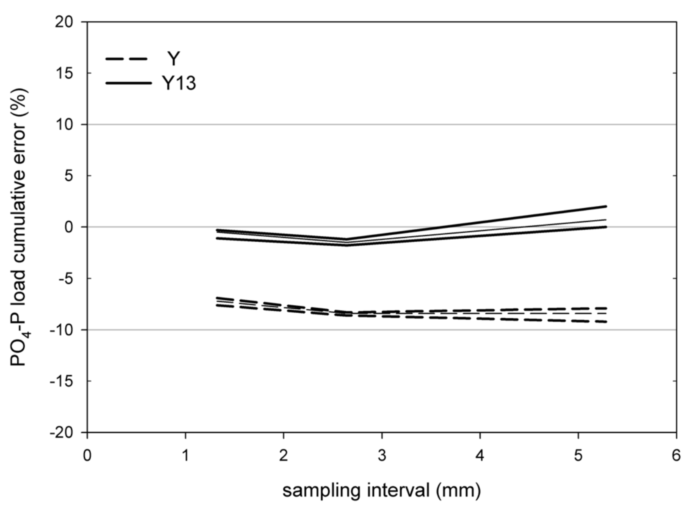

Figure 5. Cumulative error in PO_4-P load determination as a funciton of sampling interval for discrete and composite sampling; the maximum, minimum, and average percent error are presented.

Table 3. Root mean square error (RMSE), error sum of squares (errorSS), and mean absolute errors (avgABS) based on sampling interval (mm) and the number of composite samples.

Sampling Strategy		Sediment			NO_3-N			PO_4-P		
Composite (No.)	Interval (mm)	RMSE	ErrorSS	AvgABS	RMSE	ErrorSS	AvgABS	RMSE	ErrorSS	AvgABS
Watershed Y										
Discrete	1.32	12.1	2343.3	−2.7	0.01	0.39	0.00	0.02	0.00	0.00
2	1.32	12.6	2183.8	−2.8	0.03	0.39	0.01	0.02	0.00	0.00
3	1.32	12.7	2518.3	−2.8	0.03	0.39	0.01	0.02	0.00	0.00
4	1.32	12.7	2579.9	−2.8	0.01	0.39	0.00	0.02	0.00	0.00
5	1.32	11.7	2642.1	−2.6	0.02	0.37	0.00	0.02	0.00	0.00
Discrete	2.64	1.3	3221.1	0.3	0.01	0.31	0.00	0.02	0.00	0.00
2	2.64	1.4	3360.7	0.3	0.03	0.31	−0.01	0.02	0.00	0.00
3	2.64	3.6	3365.7	0.8	0.04	0.31	−0.01	0.02	0.00	0.00
4	2.64	4.3	3515.9	1.0	0.04	0.71	0.01	0.02	0.00	0.00
5	2.64	4.0	3205.9	0.9	0.03	0.28	−0.01	0.02	0.00	0.00
Discrete	5.28	7.5	4158.8	−1.7	0.08	0.23	−0.02	0.02	0.00	0.00
2	5.28	4.5	3755.5	−1.0	0.11	0.25	−0.02	0.02	0.00	0.00
3	5.28	4.8	3760.6	−1.1	0.10	0.27	−0.02	0.02	0.00	0.00
4	5.28	0.2	3081.1	0.0	0.09	0.25	−0.02	0.02	0.00	0.00
5	5.28	2.5	3210.8	0.6	0.10	0.25	−0.02	0.02	0.00	−0.01
Watershed Y13										
Discrete	1.32	19.2	4958.1	−4.7	0.58	2.48	−0.14	0.00	0.00	0.00
2	1.32	24.7	7283.0	−6.0	0.69	3.10	−0.17	0.01	0.00	0.00
3	1.32	25.8	5365.5	−6.3	0.75	3.72	−0.18	0.00	0.00	0.00
4	1.32	28.0	6152.8	−6.8	0.67	2.91	−0.16	0.00	0.00	0.00
5	1.32	18.0	4135.7	−4.4	0.88	5.73	−0.21	0.00	0.00	0.00
Discrete	2.64	4.6	10218.2	−1.1	0.39	3.49	−0.09	0.01	0.01	0.00
2	2.64	12.8	7941.9	−3.1	0.35	2.91	−0.08	0.01	0.00	0.00
3	2.64	9.4	9573.1	−2.3	0.40	3.24	−0.10	0.01	0.01	0.00
4	2.64	5.5	10663.7	−1.3	0.19	2.69	−0.05	0.01	0.01	0.00
5	2.64	1.8	6137.1	−0.4	0.52	4.12	−0.12	0.01	0.00	0.00
Discrete	5.28	56.1	39670.7	−13.6	0.31	2.71	−0.08	0.01	0.00	0.00
2	5.28	9.5	30290.7	−2.3	0.34	3.84	−0.08	0.00	0.00	0.00
3	5.28	43.6	37668.9	−10.6	0.20	2.71	−0.05	0.00	0.00	0.00
4	5.28	26.1	29201.6	−6.3	0.07	3.56	−0.02	0.00	0.00	0.00
5	5.28	18.8	30177.7	−4.6	0.37	3.03	−0.09	0.00	0.00	0.00

sampling interval. Although this option is preferred for load determination, increasing the number of composite samples reduces the knowledge of the mechanisms of within-event constituent behavior. Composite sampling is an important option because discrete strategies with small sampling intervals cannot completely sample large-volume runoff events within the common 24-bottle capacity (for example, the capacity of the 1.32 mm discrete strategy used in this study was 31.7 mm). In large runoff events, discrete strategies may produce substantial error in spite of small sampling intervals because the events are not sampled throughout their complete duration.

All of the flow-interval strategies evaluated in this study, which include sampling intervals from 1.32 to 5.28 mm, produced cumulative load errors less than ±10%; however, the effect of sampling interval on absolute event and cumulative errors was most pronounced for sediment, less for NO_3-N, and even less for PO_4-P. It is expected that sampling intervals up to 6 mm should produce similar load accuracy in other locations for constituents that vary little within runoff events based on the CV (such as PO_4-P in this study), but smaller intervals (1 to 3 mm) should be used to limit uncertainty for constituents that vary more (such as sediment in this study).

The ranges of relative cumulative error for flow-interval sampling strategies with intervals less than 6 mm are similar to those reported in an analytical study by King and Harmel (2003). In that study, flow-based sampling at 5.0 mm intervals produced a similar number of samples as time-based sampling at 180 min intervals but with much lower error ranges (−5% to 13% compared to −30% to 24%). The combined results from these two studies support a common theory previously based only on analytical data and limited field data from perennial sites, which is that flow-interval sampling better represents storm loads than time-interval strategies (Claridge, 1975; Richards and Holloway, 1987; Rekolainen et al., 1991; Shih et al., 1994; Izuno et al. 1998; Miller et al., 2000). In practical terms, it is very difficult to choose time intervals that are able to completely sample events of varying duration with adequate frequency to capture constituent concentration behavior without exceeding sampler capacity; however, it is much easier for flow-interval strategies to intensively sample throughout entire events of varying magnitude.

More research, such as the present study and a recent study by Agouridis and Edwards (2003), is needed to address the issue of adequate storm water quality characterization within resource and equipment constraints. Agouridis and Edwards (2003) developed a novel method to estimate the flow-weighted mean concentration (equivalent to the EMC) for single-peak hydrographs by selecting one discrete sample and determining its relationship to hydrologic parameters. Their method reduces analysis costs by requiring the analysis of only one sample for each event, but it was evaluated only for single-peak hydrographs. While their analytical method is well conceived, the single-bottle, flow-interval composite sampling strategy used in this study and described in Harmel et al. (2003b) remains a proven alternative. This strategy, which has been used successfully by the U.S. Geological Survey (Raymond Slade, personal communication, 2004), produces the EMC directly from one composite sample without analytical estimation. Other advantages are its ability to capture single- or multiple-peak hydrograph events and to collect a large number of relatively small-volume samples, which allows large storms to be completely sampled at a high sampling frequency, thus minimizing sampling error. This composite strategy does, however, prevent quantification of within-event concentration changes, which is a potential disadvantage. It is important to remember that sampling strategies should be designed based on specific sampling goals, as discussed for small watersheds in Harmel et al. (2003b), so any one sampling strategy will not be best suited for all projects.

The procedures and results presented in this study should apply directly to most "edge of field" and small stream sampling conditions for constituents with uniform concentrations in the flow cross-section; but whether the results can be extrapolated to differing hydrologic conditions, such as large river (basin-scale) sampling, is unknown. The results should not be applied to nutrient loads immediately below point-source inputs or to total sediment loads, especially when bed load transport is substantial, because of considerable vertical and cross-sectional concentration variability in the flow profile.

SUMMARY AND CONCLUSIONS

Most storm water quality sampling projects face the problem of balancing sampling and analysis resources with the need for accurate water quality characterization, which typically involves managing the number of samples collected without increasing sampling error. However, appropriately addressing this issue is especially difficult because the effects of sample collection techniques on uncertainty are rarely evaluated. It is hoped that the information presented in this study provides a basis for understanding uncertainty in measured storm water quality data from small agricultural watersheds. Most importantly:

- The most effective method to achieve this balance, especially for projects in which load determination is the primary goal, is composite sampling because it increases sampler capacity and creates no significant effect on sampling error.
- Intensive flow-interval sampling strategies (<6 mm) were able to sample events with a wide range of durations with adequate frequency to produce relatively low levels of uncertainty. Cumulative errors were less than $\pm$10% for all flow-interval sampling strategies evaluated (1.32 to 5.28 mm sampling intervals and discrete and composite sampling of up to five samples per bottle).
- For a given number of samples collected, flow-interval sampling better represents actual storm loads than time-interval sampling. In practical terms, it is also much easier to design flow-interval strategies that intensively sample throughout entire events of varying magnitude.
- To limit uncertainty in terms of absolute error, flow-based sampling intervals from 1 to 6 mm should be adequate for constituents that vary little within runoff events (such as nutrients from unfertilized areas), but smaller intervals (1 to 3 mm) should be used for constituents that vary more (such as sediment from highly erodible areas).

ACKNOWLEDGEMENTS

The technicians involved in this study, Larry Bartek, James Haug, Gary Hoeft, and Larry Koester, deserve credit for their valuable input, equipment design, and data collection. June Wolfe's experience in automated sampling techniques also contributed greatly to this study. Raymond Slade (Retired Hydrologist and Texas District Surface-Water Specialist, United States Geological Survey), Tom Haan (Emeritus Professor, Oklahoma State University), and Clarence Richardson (Agricultural Engineer and Laboratory Director, USDA-ARS) also deserve recognition for providing valuable insight and review of the manuscript. The Texas State Soil and Water Conservation Board provided funding for this project under its Total Maximum Daily Load (TMDL) Program.

REFERENCES

Agouridis, C. T., and D. R. Edwards. 2003. The development of relationships between constituent concentrations and generic hydrologic variables. *Trans. ASAE* 46(2): 245-256.

Claridge, G. G. C. 1975. Automated system for collecting water samples in proportion to stream flow rate. *New Zealand J. Sci.* 18: 289-296.

Guo, Y., M. Momcilo, and D. Misganaw. 2002. Uncertainty of nitrate-N load computations for agricultural watersheds. *Water Resources Res.* 38(10): 1185-1196.

Haan, C. T. 2002. *Statistical Methods in Hydrology.* 2nd ed. Ames, Iowa: Iowa State University Press.

Haggard, B. E., T. S. Soerens, W. R. Green, and R. P. Richards. 2003. Using regression methods to estimate stream phosphorus loads at the Illinois River, Arkansas. *Applied Eng. in Agric*. 19(2): 187-194.

Harmel, R. D., K. W. King, J. E. Wolfe, and H. A. Torbert. 2002. Minimum flow considerations for automated storm sampling on small watersheds. *Texas J. Sci*. 54(2): 177-188.

Harmel, R. D., K. W. King, C. W. Richardson, and J. R. Williams. 2003a. Long-term precipitation analyses for the central Texas Blackland Prairie. *Trans. ASAE* 46(5): 1381-1388.

Harmel, R. D., K. W. King, and R. Slade. 2003b. Automated storm water sampling on small watersheds. *Applied Eng. in Agric.* 19(6): 667-674.

Helsel, D. R., and R. M. Hirsch. 1993. *Statistical Methods in Water Resources.* New York, N.Y.: Elsevier.

Hession, W. C., D. E. Storm, C. T. Haan, K. H. Reckhow, and M. D. Smolen. 1996. Risk analysis of total maximum daily loads in an uncertain environment using EUTROMOD. *J. Lake and Reservoir Mgmt.* 12(3): 331-347.

Izuno, F. T., R. W. Rice, R. M. Garcia, L. T. Capone, and D. Downey. 1998. Time versus flow composite water

King, K. W., and R. D. Harmel. 2003. Considerations in selecting a water quality sampling strategy. *Trans. ASAE* 46(1): 63-73.
King, K. W., and R. D. Harmel. 2004. Comparison of time-based sampling strategies to determine nitrogen loading in plot-scale runoff. *Trans. ASAE* 47(5): 1457-1463.
Martin, G. R., J. L. Smoot, and K. D. White. 1992. A comparison of surface-grab and cross-sectionally integrated stream-water-quality sampling methods. *Water Environ. Res*. 64(7): 866-876.
McFarland, A., and L. Hauck. 2001. Strategies for monitoring nonpoint-source runoff. TIAER Report No. 0115. Stephenville, Texas: Texas Institute for Applied Environmental Research.
Miller, P. S., B. A. Engel, and R. H. Mohtar. 2000. Sampling theory and mass load estimation from watershed water quality data. ASAE Paper No. 003050. St. Joseph, Mich.: ASAE.
Minitab. 2000. *Minitab Statistical Software User's Guide 2: Data Analysis and Quality Tools, Release 13.* State College, Pa.: Minitab, Inc.
Parsons, D. A. 1954. Coshocton-type runoff samplers: Laboratory investigations. SCS-TP-124. Washington, D.C.: USDA-SCS.
Rekolainen, S., M. Posch, J. Kamari, and P. Ekholm. 1991. Evaluation of the accuracy and precision of annual phosphorus load estimates from two agricultural basins in Finland. *J. Hydrology* 128: 237-255.
Richards, R. P., and J. Holloway. 1987. Monte Carlo studies of sampling strategies for estimating tributary loads. *Water Resources Res.* 23(10): 1939-1948.
Robertson, D. M., and E. D. Roerish. 1999. Influence of various water quality sampling strategies on load estimates for small streams. *Water Resources Res*. 35(12): 3747-3759.
Shih, G., W. Abtew, and J. Obeysekera. 1994. Accuracy of nutrient runoff load calculations using time-composite sampling. *Trans. ASAE* 37(2): 419-429.
Stone, K. C., P. G. Hunt, J. M. Novak, M. H. Johnson, and D. W. Watts. 2000. Flow-proportional, time-composited, and grab sample estimation of nitrogen export from an eastern Coastal Plain watershed. *Trans. ASAE* 43(2): 281-290.
Tate, K. W., R. A. Dahlgren, M. J. Singer, B. Allen-Diaz, and E. R. Atwill. 1999. Timing, frequency of sampling affect accuracy of water-quality monitoring. *California Agric.* 53(6): 44-49.
USEPA. 1995. National water quality inventory 1994 report to Congress. USEPA 841-R-95-005. Washington, D.C.: U.S. EPA, Office of Water.
USEPA. 2000. National water quality inventory 1998 report to Congress. USEPA 841-R-00-001. Washington, D.C.: U.S. EPA, Office of Water.
Wells, F. C., W. J. Gibbons, and M. E. Dorsey. 1990. Guidelines for collection and field analysis of water-quality samples from streams in Texas. USGS Open-File Report 90-127. Washington, D.C.: United States Geological Survey.