INFLUENCE OF DIFFERENT TEMPORAL SAMPLING STRATEGIES ON ESTIMATING LOADS AND MAXIMUM CONCENTRATIONS IN SMALL STREAMS

DALE M. ROBERTSON AND KEVIN D. RICHARDS

Dale M. Robertson is a research hydrologist with the U.S. Geological Survey,WRD (8505 Research Way, Middleton, WI 53562). His research deals with examining the effects of different sampling strategies, estimating loads and concentrations over large geographic areas, and water-quality and numerical modeling of lakes.

Kevin D. Richards is a physical scientist with the U.S. Geological Survey, WRD (8505 Research Way, Middleton, WI 53562). His activities deal with data collection and database management of water-quality studies of streams.

ABSTRACT

Various temporal sampling strategies are used to monitor small streams to estimate annual loads and maximum concentrations of specific constituents. Extensive water-quality data from eight small streams in Wisconsin were systematically subsampled to simulate and evaluate typically used strategies. These subsets of data were then used with daily average streamflow to estimate annual loads using the regression method (the most commonly used method for streams with limited water-quality data). For each strategy, the accuracy and precision of estimated loads and maximum concentrations of phosphorus were evaluated by comparison with "true" annual loads (computed with the integration method) and maximum concentrations, both estimated using all available data for each site during the period of interest.

The most cost-effective sampling strategy depends on whether loads or maximum concentrations are examined and the duration of the study. For 1-year studies, fixed-period monthly sampling supplemented by storm chasing was most cost-effective for estimating annual loads, even though loads were overestimated by 25-50%. For 2 to 3-year studies, fixed-period, semi-monthly sampling provided not only the least biased but also the most precise loads. Any strategy that collects samples on the rising limbs of hydrographs was best for detecting maximum concentrations regardless of the duration of the study. The high-flow samples, especially if consistently collected during rising stage, do not represent the average daily concentration and therefore result in imprecise, over-estimated annual loads if used in the regression.

1. INTRODUCTION

Various temporal strategies have been used to collect samples to describe water quality in small streams. These range from strategies trying to describe all of the changes in concentration to simple fixed-period strategies, such as monthly. Although costs of implementing different strategies vary, the monitoring programs often have the same common goals, such as describing maximum concentrations occurring in a stream or estimating annual loads. With intense sampling, detection of maximum concentrations and estimation of annual loads (with an integration approach) is straight forward and accurate. However, with less intense sampling, different assumptions and approaches must be used to estimate maximum concentrations and annual loads and may reduce the accuracy in these estimates.

In small streams, concentrations of many constituents are directly related to streamflow. For example, constituents associated with sediments generally increase very rapidly during increasing flow, peak prior to maximum flow, and then decrease more slowly during decreasing flow (hysteresis). Changes in streamflow and total phosphorus concentrations during a high-flow event in Bower Creek, Wisconsin are shown in Fig. 1. Various strategies have been used to collect samples to take advantage of these systematic changes in concentration. The most extensive sampling design typically includes fixed-period, manually collected samples during times of stable flow when concentrations are also expected to be stable, supplemented with many samples collected with automatic samplers during changing flows and expected changing concentrations. This strategy is used by the Nonpoint Program of the U.S. Geological Survey (USGS) [Graczyk et al., 1993] and typically results in 100-200 samples per year per site for small streams, streams less than ~100 km^2. Another strategy, more typically used by long-term monitoring programs, is to manually collect samples at a fixed interval. The Wisconsin Department of Natural Resources (WDNR) collects samples monthly [Tiegs, 1986] and the Illinois Environmental Protection Agency collects samples every 6 weeks [Illinois Environmental Protection Agency, 1996]. Other sampling strategies are between these two extremes and have fixed-period sampling supplemented with a few high-flow samples. The USGS's National Water-Quality Assessment (NAWQA) Program [Hirsch et al., 1988], collects fixed-period monthly samples supplemented by four to eight manually collected high-flow samples per year for approximately 2.5 years or about 18 samples per year for both large rivers and small streams [Gilliom et al., 1995].

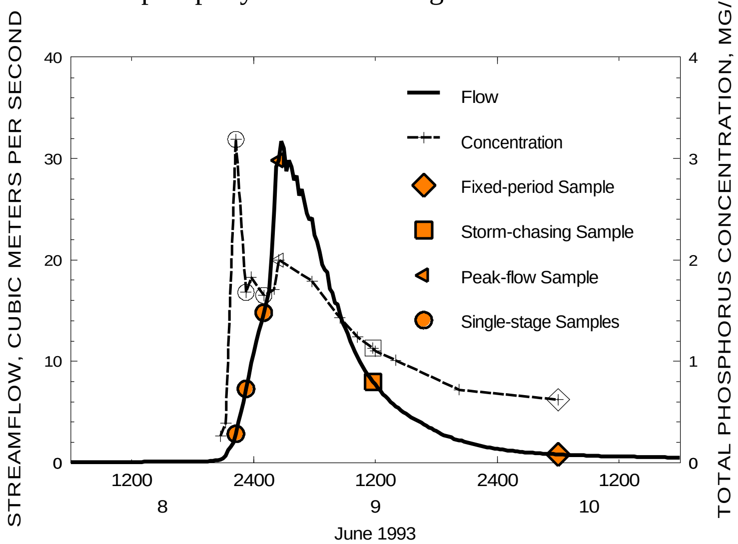

Figure 1 Changes in total phosphorus concentration during a high-flow event in a small stream in 1993 (Bower Creek, Wisconsin). Samples collected with various temporal strategies are identified with respect to flow (solid symbols) and concentration (open symbols).

Data assembled using these strategies are typically used to describe the maximum concentrations and the total transport (load) of specific constituents in streams. Various approaches have been used to quantify the annual load in small streams. In most of the more accurate approaches, it is assumed that discrete water-quality samples are collected and that continuous or at least daily average streamflow records are available or can be estimated. The most common approach is to estimate loads by means of continuous (at least daily) concentration and streamflow traces. Loads are then estimated by multiplying the continuous concentration trace by the continuous streamflow trace. A continuous concentration trace can be developed by either of two approaches: the integration method and regression (rating curve) method.

In the integration method, constituent concentrations are plotted through time, and hydrologic judgement is used to extrapolate between measured concentrations [Porterfield, 1972]. This method is generally considered to be the most accurate method to estimate loading at all time scales if sufficient data are collected to describe the changes in water quality, especially if samples are collected throughout the largest high-flow events during the period. It is difficult, however, to place confidence limits on loads estimated with this approach. For accurate load estimations, "sufficient data" often means that many samples be collected to reflect the variability in water quality; thus, this is the most expensive approach.

The regression method typically uses a relation found between concentration (or load) and daily average flow (and other independent variables) to estimate daily concentrations (or loads) of the constituent. This method began as simple linear relations between concentration (or load) and flow but has been modified to account for nonlinearities, seasonal and long-term variability, censored data, biases associated with using logarithmic transformations, and serial correlations in the residuals of the analyses [Cohn, 1995]. This approach has come into widespread use because it requires less data than the integration approach, produces estimates for periods beyond when concentration data were collected, and enables confidence limits to be placed on the estimates. The regression method is often used with very small data sets that have been assembled over several years. With this method, one typically uses daily average streamflow to estimate daily average concentrations (or loads) because this is the resolution of most streamflow databases. Therefore, each concentration and streamflow combination (and any other independent variable) used in the regression is assumed to be representative of the average daily conditions when the data were collected. Therefore, high-flow samples should be randomly collected throughout the day during high-flow events. This type of regression approach is usually considered a "big river" approach because it is based on the assumption that samples represent the daily average concentration and it estimates changes in concentration (or loads) on a daily time step. However, this approach is commonly used to estimate loads in small streams in which concentrations change at time scales much less than a day [Walker, 1996].

The accuracy of estimating annual loads and detecting maximum concentrations with very limited data has been evaluated in only a few cases. Walling and Webb [1981] used hourly suspended sediment data, derived from continuous turbidity measurements, collected over seven years to estimate "true" annual sediment loads for a small stream in the United Kingdom using the integration approach. These data were then used to determine the accuracy and precision of the regression approach using subsamples of these data chosen based on fixed intervals (1 to 14 days) and a fixed interval (7 days) supplemented with additional random samples collected during high flows. They found that the annual loads estimated by use of the regression approach were consistently underestimated by 23 to 83%. Walling and Webb [1988] obtained similar results for other small streams. Dolan et al. [1981] and Preston et al. [1989] evaluated the regression approach to estimate annual loads for two large rivers. Virtually daily data were used to estimate "true" loads using the integration approach. These values were compared with estimates made by the regression approach using daily average streamflow and randomly selected subsets of the water-quality data (12 samples to represent monthly, 4 to represent quarterly, etc.) or stratified-random data sets (randomly monthly plus randomly during high flows). They found average biases and standard deviation in the errors in annual loads to be less than 10% when using about 24 samples a year and only minor improvement when substituting 12 monthly samples plus 12 event samples for the 24 semimonthly samples.

Few studies have examined how specific temporal sampling strategies affect estimated annual loads; however, one such study was done by Robertson and Roerish [1999]. In that study, extensive water-quality data from eight small streams in Wisconsin were systematically subsampled to simulate and evaluate typically used temporal sampling strategies. In this paper, we summarize those results and expand the analysis to examine the effects on maximum concentrations and project costs. The ultimate goal of this work is to determine the most cost-effective monitoring program for small streams as a function of the duration of the study.

2. STUDY SITES

The eight small stream sites are located in agricultural areas of the southern half of Wisconsin and have drainage areas that range in size from 14 to 110 km^2 (Table 1). Each site was instrumented to continuously record water levels and compute flow by use of a stage-discharge relation. Water samples were collected manually at fixed intervals (~ every 2 weeks from March through October and monthly in other months) by use of the equal-width increment (EWI) method described by Guy and Norman [1970] and throughout high-flow events by use of stage-activated, refrigerated, automatic samplers [Graczyk et al., 1993]. At each site, 93 to 195 water samples were collected each year (Table 1; approximately 20 fixed-period samples and usually 6 to 10 samples during each of about 10 to 20 storms, annually) and analyzed for total phosphorus (TP). All chemical analyses were done by the Wisconsin State Laboratory of Hygiene in accordance with the guidelines of the U.S. Environmental Protection Agency [Wisconsin State Laboratory of Hygiene, 1993]. A detailed summary of collection procedures and quality assurance and quality control is given by Graczyk et al. [1993]. Data-collection periods for these sites vary, but each was sampled during 1992-94. Included in this period was a relatively dry year (1992), a relatively wet year (1993), and a relatively normal year (1994).

Table 1. Characteristics of eight streams used in the study

Site Name	U.S. Geological Survey Station Number	Drainage Area, km^2	Average Number of Samples per Year
Bower Creek	04085119	38.3	195
Brewery Creek	05406470	27.2	100
Eagle Creek	05378185	37.0	100
Garfoot Creek	05406491	14.0	93
Joos Valley Creek	05378183	15.3	107
Kuenster Creek	054134435	24.9	98
Otter Creek	040857005	24.6	172
Rattlesnake Creek	05413449	109.8	117

3. METHODS

3.1. Sampling Strategies

To simulate data collected using sampling strategies of typical monitoring programs, the TP time series of each stream was systematically subsampled. Ten sampling strategies were simulated. Three were fixed-period strategies (semimonthly, monthly, and every 6 weeks). Seven involved fixed-period plus high-flow samples (semimonthly plus storm-chasing samples, semimonthly plus single-stage samples, monthly plus storm-chasing samples, monthly plus peak-flow samples, monthly plus single-stage samples, 6-week plus storm-chasing samples and 6-week plus single-stage samples). Three study periods were examined to simulate the duration of typical monitoring studies: 1 year (dry, 1992; and wet, 1993), 2 years (1992-93), and 3 years (1992-94). Only six months of data were used for 1994 to simulate the typical 2.5-year sampling period of NAWQA.

3.2. Subsampling Data

Fixed-period monthly sampling was simulated by using the sample collected at the most frequently sampled time of the month in the entire period for each site. Semimonthly sampling was simulated by using the monthly sample plus a second sample collected ~15 days away from the monthly sample. Six-week sampling was simulated by using every third sample in the semimonthly data set, unless a semimonthly sample was missing, such as often occurred during winter. If more than one sample was collected on any of these dates, the sample collected closest to 11:00 a.m. was chosen (midmorning was when most manual samples were collected).

After testing various protocols, the following strategy was selected to subsample high-flow events in an unbiased manner and provide ~ 6 to 9 high-flow samples per year. The first step was to determine instantaneous flow thresholds for a high-flow event that become more stringent in wet years with many events and less stringent during dry years:

A. Base threshold: Compute the flow at the 99th percentile of flow from existing daily average flow data, i.e., the daily average flow that is exceeded only 1% of the time (TB).
B. More stringent threshold: TM1 = TB * 2.
C. Most stringent threshold: TM2 = TB * 4.
D. Less stringent threshold: TL1 = TB * 0.75.
E. Least stringent threshold: TL2 = TB * 0.50.

The second step was to determine when high-flow thresholds (based on historical daily average flows) were reached by examining the instantaneous (15-minute) flow data stored in the USGS Automated Data-Processing System (ADAPS) within the National Water Information System (NWIS) [U.S. Geological Survey, 1998] and selecting the samples appropriate for the various sampling strategies. To decide which samples would have been collected by a storm-chasing crew, a sample was chosen two hours (response time) after the instantaneous flow surpassed the high-flow thresholds between 6:00 a.m. and 4:00 p.m. (typical work hours). Future flow data were not considered because a sampling crew would not have this information. Only one sample was collected in any high-flow event. Once a high-flow sample was chosen in a month, the threshold became more stringent and set to TM1. If a second sample was selected in a given month, the threshold was raised to TM2. If two full months passed without selection of a high-flow sample, the threshold was relaxed to TL1 and if four full months passed, it was further relaxed to TL2. Once the first high-flow sample was collected, the threshold was always set to TM1 and to TB the following month. This approach usually resulted in ~6 high-flow samples per year. To simulate samples collected at peak flow, the sample collected nearest peak flow, regardless of the time of day, was chosen for each event sampled by the hypothetical storm-chasing crew.

Another strategy to collect high-flow samples is by the use of single-stage samplers. Single-stage samplers, also referred to as siphon samplers, automatically collect a sample when the water level of a stream passes a specified elevation [Edwards and Glysson, 1988]. For this strategy, samplers were simulated to be set at heights equivalent to flows of TB, TM1, and TM2. Samples were then selected closest to when the flow first exceeded these thresholds regardless of the time of day; therefore, samples were only collected during increasing flow. Only one sample per month was selected for each threshold, and a maximum of three were selected for each threshold per year. A maximum of nine single-stage samples were chosen each year. Samples selected using these strategies are shown in Fig. 1.

3.3. Load Computation

Daily, event, and annual TP loads were previously computed for each site by use of the integration method described by Porterfield [1972] and published in annual USGS reports and stored in the NWIS database.

Annual loads (calculated by summing daily loads) were estimated by a regression approach using the Estimator program [Cohn et al., 1989]. Estimated daily loads (L) were computed based on relations between constituent load and two variables: streamflow (Q) and time of the year (T, in radians). The general form of the model was

(1) $$\log (L) = a + b [\log (Q) - c] + d [\log (Q) - c]^2 + e [\sin (T)] + f [\cos (T)].$$

Values for the coefficients (a, b, c, d, e, and f) in Eq. 1 were computed for each site and for each time period by the use of multiple regression analyses between daily loads (daily average Q's multiplied by instantaneous measured concentrations) and daily average Q and T. In each regression, only terms that were significant at $P < 0.05$ were included. Because a logarithmic transformation was used in Eq. 1, daily loads were adjusted to account for a retransformation bias by use of the minimum variance unbiased estimate procedure (see Cohn et al. [1989] for a complete discussion). To combine the errors for the eight streams and allow comparison among streams and years, all annual loads were converted to yields (load per unit area) and errors were normalized as percentage of the "true" annual yield as determined by the integration method or as a percentage of the "true" maximum concentration from all of the data for the defined periods.

3.4. Evaluation Criteria

Estimated annual loads and maximum concentrations from the various strategies were compared with the "true" loads (computed by the integration approach) and maximum concentrations (using all of the data for the defined period) to describe the errors. Overall errors are dependent on the magnitude of two components (accuracy and precision), which have been combined into one overall estimated error called the normalized mean square error (MSE, in Eq. 2) [Preston et al., 1989]. The accuracy or bias represents the average or median difference between the estimates and the "true" values. The precision represents the measure of the spread or variance of the errors (s^2; computed as the standard deviation of the errors squared).

(2) $$MSE = Bias^2 + s^2$$

3.5. Costs

The total costs of implementing the various strategies were subdivided into setup (for single-stage samplers), analytical, personnel (collection costs and preparation and handling costs), and mileage costs. Costs were computed for two different sets of constituents to be examined: limited set (field parameters, phosphorus, and suspended sediments - $30 per sample) and an extensive set (field parameters, nutrients, major ions, and pesticides - $510 per sample). In determining the costs it was assumed that: eight sites were to be monitored, each site was ~50 miles from the office and 50 miles from one another. Labor costs were $15 per hour. Three single-stage samplers were installed at each site ($270 total; $150 equipment costs and 8 person hours to install). Mileage costs were $0.32 per mile.

The number of samples that can be collected in one 10-hour day varies with each sampling strategy. Fixed-period sampling can take advantage of a planned travel routine; therefore, it was assumed that 4 sites could be visited per day (200 total miles per trip or 50 miles per site) for a limited set of constituents and only 2 sites per day (150 total miles per trip or 75 miles per site) for an extensive set. Storm-chasing requires monitoring real-time transmitted hydrographs; therefore, it was assumed that only 2 sites could be visited per day (75 miles per site) for limited analyses and 1.5 sites per day (average of 125 total miles per trip or 88 miles per site) for extensive analyses. Single-stage samples do not usually require an instant response and often have several samples collected during one event; therefore, it was assumed that 6 samples could be collected per day (assumed 50 miles per sample) for limited analyses and 3 samples per day (assumed 50 miles per sample) for extensive analyses. Samples collected at peak flow require on-location monitoring of the stream; therefore, it was assumed that only 1 site could be visited per day (100 miles per site) for limited and extensive analyses. Preparation and processing of samples depends on the number of samples collected in a day; therefore, it was assumed in an 8-hour day that 8 fixed-period samples, 4 storm or peak-flow samples, or 12 single-stage samples could be processed.

4. RESULTS

4.1 "True" Estimations

Annual yields (computed with the integration method) and maximum concentrations for each of the eight sites were computed using all of the data collected at each site for the specified period (Table 2). Annual TP yields ranged from 19 to 678 kg/km^2 (average annual yields ranged from 61 to 315 kg/km^2). For each stream, the highest annual yield occurred in 1993 (wet year), and, in most cases, the lowest annual yield occurred in 1992 (dry year). Annual maximum concentrations ranged from 1.4 to 25.0 mg/L (for the entire period, maximum concentrations ranged from 4.1 to 25.0 mg/L). Most of the maximum concentrations during the period again occurred in 1993.

Table 2. Total annual phosphorus yields (load per unit area) and maximum concentrations for the eight streams studied. All loads were estimated by use of the integration method

	Total Phosphorus Yield, kg/km^2				Maximum Concentration, mg/L			
Site Name	1992	1993	1994	Average	1992	1993	1994	Overall
Bower Creek	177	248	104	177	2.9	3.2	4.1	4.1
Brewery Creek	19	181	56	85	5.7	25.0	11.7	25.0
Eagle Creek	82	217	144	148	5.4	13.6	19.8	19.8
Garfoot Creek	70	209	69	116	5.1	7.4	2.2	7.4
Joos Valley Creek	103	189	149	147	9.7	15.4	14.1	15.4
Kuenster Creek	52	678	166	299	9.0	18.4	8.3	18.4
Otter Creek	35	94	55	61	1.4	6.5	3.3	6.5
Rattlesnake Creek	131	632	182	315	8.4	10.9	6.5	10.9

4.2. Comparison with "True" Values – Error Analysis

The estimated annual yields and maximum concentrations from the various strategies were compared with the "true" yields and maximum concentrations and the errors are described below in terms of biases, variances, and overall errors (MSE, in Eq. 2).

4.2.1 Annual load estimates

The effects of the various sampling strategies on the estimation of annual loads were examined for 1-, 2-, and 3-year study durations. Median biases in the annual yields estimated using each sampling strategy are shown in Fig. 2 and summarized in Table 3. Median biases were used to minimize the effects of a few outliers. Similar results were found for all study durations. Median biases ranged from almost no bias for fixed-period sampling strategies to greater than 100% for strategies with additional high-flow samples, especially when collected during increasing flow with single-stage samplers or near peak flow. Almost all biases were positive, indicating that estimated yields were greater than "true" yields. In general, examining average (rather than median) biases resulted in the same conclusions. When all data for each site were used in the regressions, there was a positive bias >50%.

Table 3. Summary of errors from the 3-year study duration for the eight streams studied

Sampling Strategy	Bias	Variance	Mean Square Error	Root-Mean-Square Error	Absolute Error
			Yield (based on median errors)		
All Data	55	4,700	7,800	88	55
Semimonthly	**10**	**1,700**	**1,800**	**43**	**33**
Semimonthly Plus Storm Chasing	46	7,200	9,300	96	46
Semimonthly Plus Single Stage	89	120,000	128,000	357	89
Monthly	**-3**	22,300	22,300	149	**32**
Monthly Plus Storm Chasing	51	5,400	8,000	90	51
Monthly Plus Peak Flow	92	17,200	25,600	160	92
Monthly Plus Single Stage	81	45,900	52,400	229	81
6-Week	**9**	368,000	368,000	606	41
6-Week Plus Storm Chasing	66	11,900	16,300	128	66
6-Week Plus Single Stage	114	139,000	152,000	390	114
			Maximum Concentration (based on mean errors)		
Semimonthly	-73	**68**	5,410	74	73
Semimonthly Plus Storm Chasing	-51	450	3,100	56	51
Semimonthly Plus Single Stage	**-13**	337	**500**	**22**	**13**
Monthly	-76	109	5,940	77	76
Monthly Plus Storm Chasing	-52	474	3,190	56	52
Monthly Plus Peak Flow	-30	593	1,480	38	30
Monthly Plus Single Stage	**-21**	620	**1,060**	**33**	**21**
6-Week	-77	**91**	6,000	77	77
6-Week Plus Storm Chasing	-52	473	3,160	56	52
6-Week Plus Single Stage	**-13**	337	**500**	**22**	**13**

Errors are in percentage of the "true" yield computed by use of the integration method or maximum concentration with all available data. The most effective sampling strategies, based on the lowest error statistic, are in **bold**.

A very small bias does not necessarily mean that a sampling strategy will produce small errors in annual loads if the variance in estimated loads is very large. Therefore, the variances in the errors were examined. For the 3-year period (fourth bar in each group in Fig. 2; 6-week strategies were examined only for the 3-year period), the variances ranged from 1,700 to 368,000. The variance in the errors was lowest for semimonthly sampling, followed by monthly sampling plus storm chasing, followed by those for semimonthly sampling plus storm chasing. High-flow samples consistently collected prior to (single-stage samples) or near peak flow increased the variance from that estimated for storm chasing. Different study durations produced very different variances compared to those found for the 3-year period; but, generally, variances were consistently lowest for monthly or semimonthly sampling supplemented by storm chasing and monthly and semimonthly sampling for longer

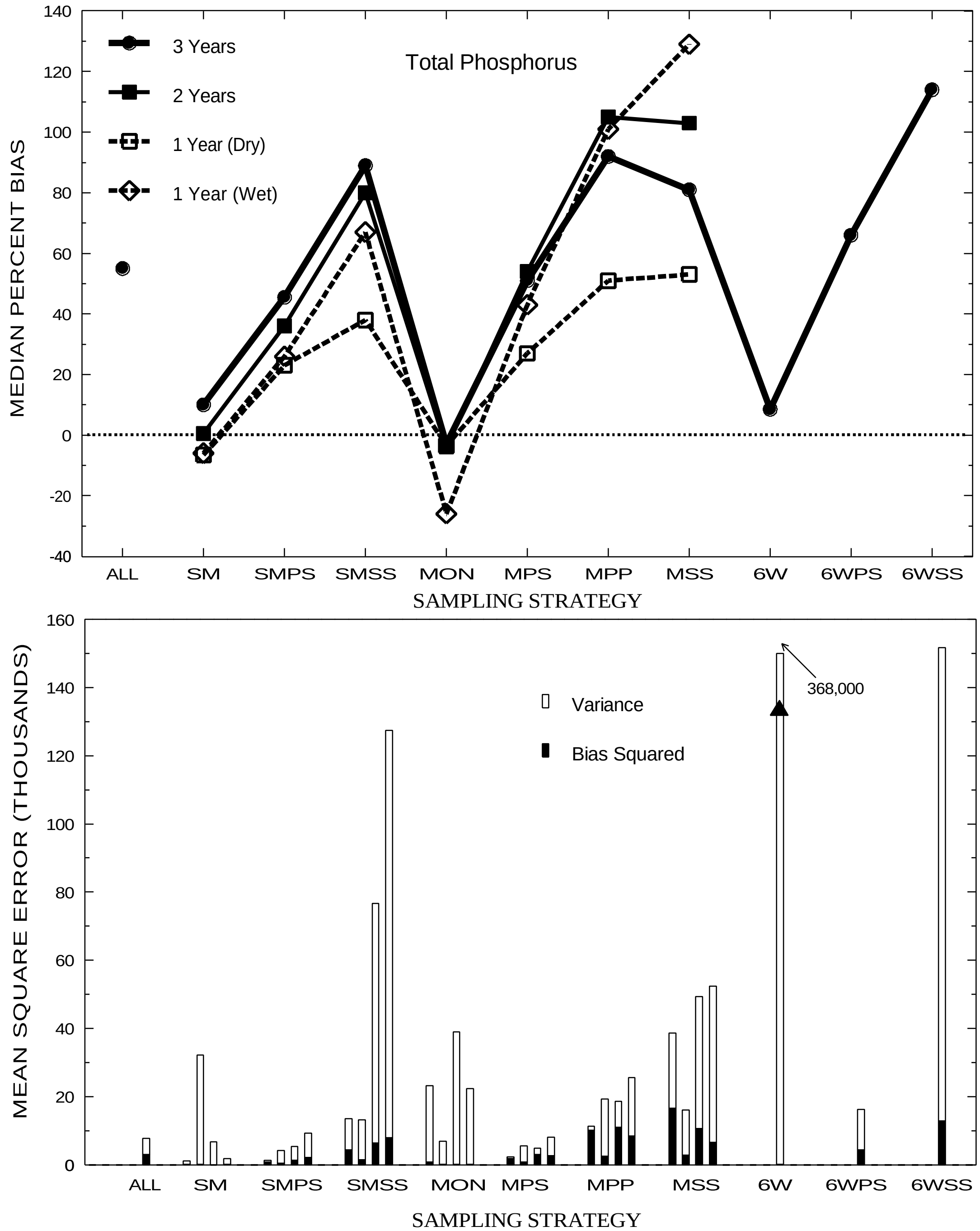

Figure 2. Biases, in percent, and mean square errors, in percent squared (percent difference from "true" yield estimated by the integration method), for total phosphorus for various sampling strategies for the eight studied streams. Each group of four bars is given in the following order: 1 year (wet), 1 year (dry), 2 years, and 3 years. The all-data and 6-week strategies were used for the 3-year periods only. [All, all data; SM, semimonthly; SMPS, semimonthly plus storm chasing; SMSS, semimonthly plus single stage; MON, monthly; MPS, monthly plus storm chasing; MPP, monthly plus peak flow; MSS, monthly plus single stage; 6W, 6-week; 6WPS, 6-week plus storm chasing; 6WSS, 6-week plus single stage.]

periods. The magnitude of the variance did not appear to have any consistent relation to the duration of the study. Sometimes 1-year studies had small errors (such as for semimonthly sampling and monthly sampling with storm chasing) and sometimes had large errors (such as for monthly sampling). In many cases, combining 2 or 3 years of samples resulted in variances larger than for each of the individual years, especially if high-flow samples were included. The variances of the errors when all of the data were used in the regressions were consistently small.

To evaluate the overall errors of the various sampling strategies, the biases and variances were combined into a MSE using Eq. 2 (Fig. 2 and Table 3). For the 3-year period, MSE's ranged from 1,800 to 368,000. To evaluate the MSE's in terms of an approximate percent error, the root-mean-square error was computed (Table 3). The lowest overall errors then equate to ~45% of the "true" annual yields. MSE's were lowest for fixed-period, semimonthly sampling. Monthly sampling plus storm chasing and semimonthly sampling plus storm chasing were in the top four strategies. Different study durations once again produced very different MSE's compared to those found for the 3-year period, primarily because the variances of the errors were generally much larger than the biases squared. MSE's were consistently lowest for monthly and semimonthly samples supplemented with storm chasing and for fixed-period monthly and semimonthly sampling for longer periods; thus, these appear to be the most effective overall sampling strategies. MSE's when all of the data were used in the regression were consistently a little larger than for these four strategies.

Another way to determine which strategies provided the best annual load estimates is to compare the median absolute errors (MAE's) (Table 3). This approach removes the importance of the sign of the error and also removes the sensitivity to outliers. For the 3-year period, the MAE's ranged from 32 to 114%. Fixed-period monthly and semimonthly sampling resulted in the smallest errors; additional high-flow samples always increased the MAE's. The largest errors generally resulted from the addition of peak-flow and single-stage samples.

4.2.2. Maximum concentrations

The mean biases in maximum concentrations estimated using each sampling strategy are shown in Fig. 3 and summarized in Table 3. Mean biases were used because all of the biases were negative and normally distributed. Very similar results were found for all study durations. Mean biases ranged from 10 to 30% for the sampling strategies that included samples consistently collected during rising or near peak flow to 60 to 80% for fixed period strategies. Biases were generally larger during wet years (first bar in each group in Fig. 3), when concentrations in the streams were higher than in dry years (second bar in Fig. 3).

For the 3-year study length (fourth bar in each group in Fig. 3), the variances ranged from 68 to 620. The variances in the errors were lowest for semimonthly and 6-week sampling, followed by monthly sampling. High-flow samples always increased the variance in the errors. Different study durations produced different variances compared to those found for the 3-year period; but, in general, variances were consistently lowest for fixed period sampling. The importance of the variances of the errors was very small compared to that of the biases (Fig. 3).

The biases and variances were combined into MSE's to access the overall errors. In this case, the errors associated with the bias were much more important than those associated with the variance. For the 3-year period, the MSE's ranged from 500 to 6,000. The lowest overall errors then equate to 22% of the "true" annual maximum concentrations and a mean absolute error (MNAE) of 13%. In general, MSE's and MNAE's were consistently lowest for any strategy that had samples collected during rising or near peak flows regardless of the length between fixed-period sampling. MSE's for all strategies that did not include sampling before or near peak flow were very large. The MSE's generally decreased with the duration of the study, although errors during dry years were relatively small. Because all biases were negative, MNAE's were identical to the absolute value of the bias.

4.3. Costs of Implementing Strategies

The average annual costs of implementing the various strategies are shown in Fig. 4 for a 3-year study for both a limited set and an extensive set of constituents. Note that these costs do not include the costs of monitoring streamflow. Similar results were found for the other study durations. For both sets, the lowest costs would be

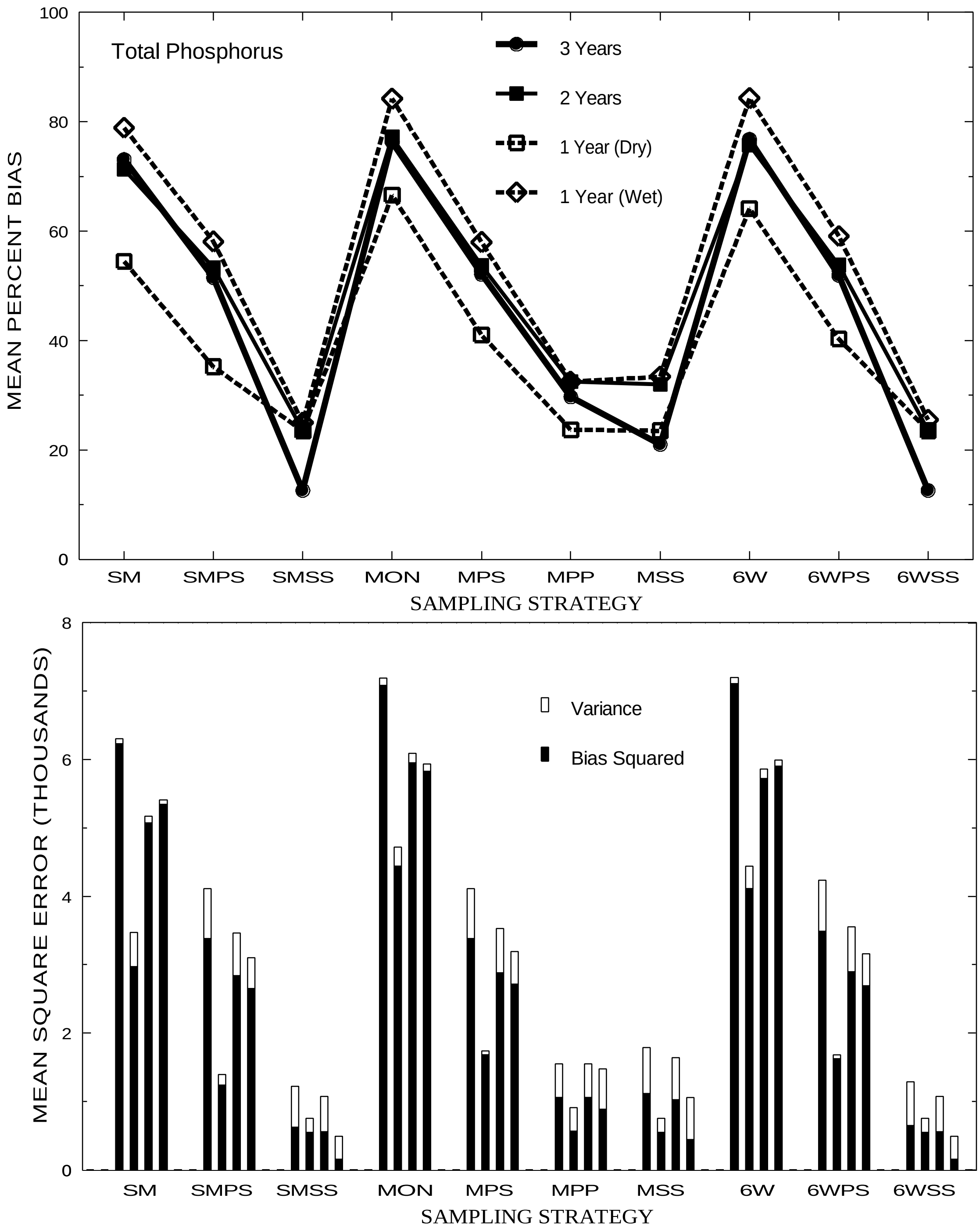

Figure 3. Biases, in percent, and mean square errors, in percent squared (percent difference from the "true" maximum concentration in the all of the data of the specified period), for total phosphorus for various sampling strategies for the eight studied streams. Each group of four bars is given in the following order: 1 year (wet), 1 year (dry), 2 years, and 3 years. [All, all data; SM, semimonthly; SMPS, semimonthly plus storm chasing; SMSS, semimonthly plus single stage; MON, monthly; MPS, monthly plus storm chasing; MPP, monthly plus peak flow; MSS, monthly plus single stage; 6W, 6-week; 6WPS, 6-week plus storm chasing; 6WSS, 6-week plus single stage.]

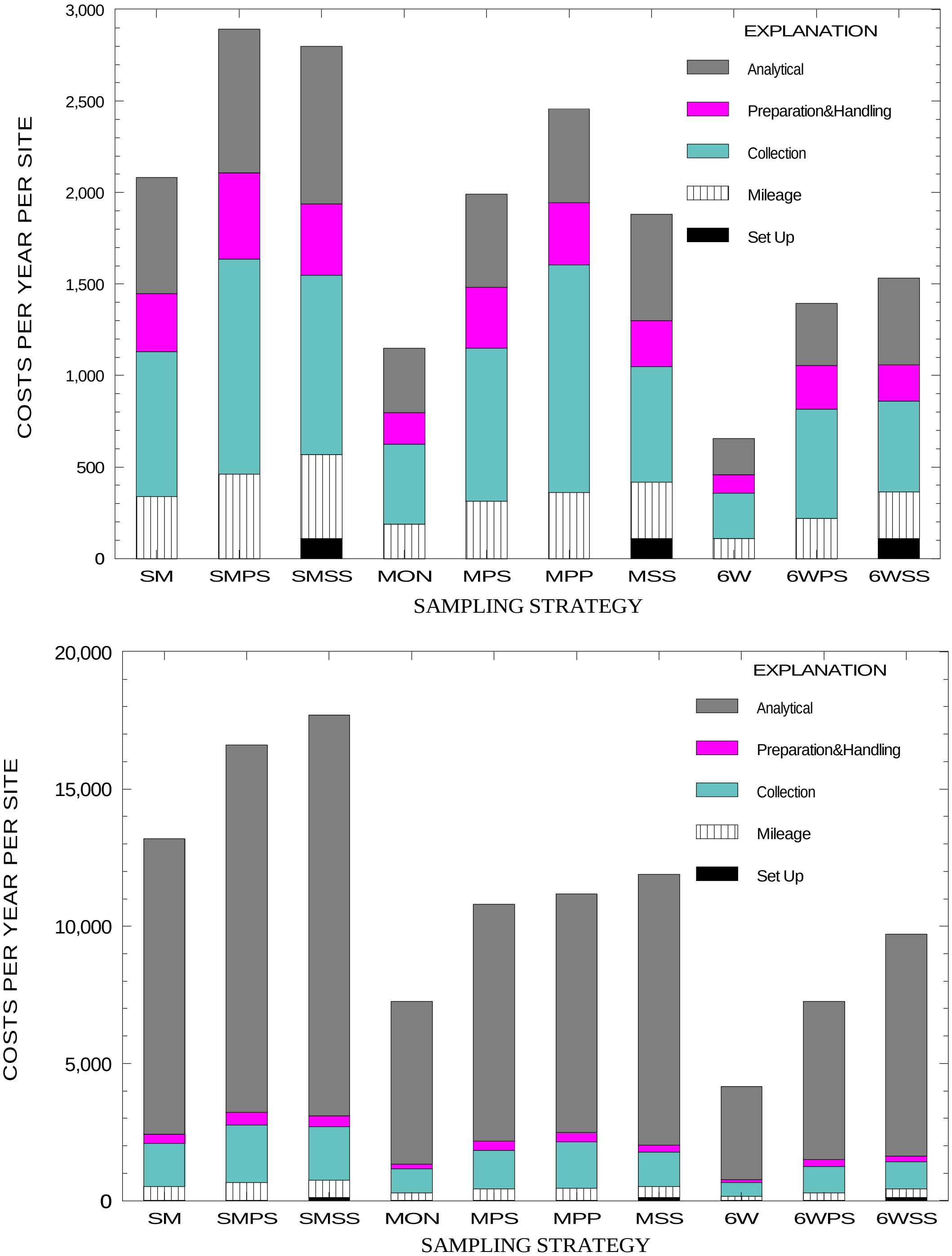

Figure 4. Average annual costs per site associated with collecting data for different temporal sampling strategies for a limited set (field parameters, phosphorus, and suspended sediments - $30 per sample) and an extensive set (field parameters, nutrients, major ions, and pesticides - $510 per sample) of constituents for a 3-year study. Other costs are described in the text. [SM, semimonthly; SMPS, semimonthly plus storm chasing; SMSS, semimonthly plus single stage; MON, monthly; MPS, monthly plus storm chasing; MPP, monthly plus peak flow; MSS, monthly plus single stage; 6W, 6-week; 6WPS, 6-week plus storm chasing; 6WSS, 6-week plus single stage.]

incurred using the fixed-period strategies (6-week followed by monthly); however, semimonthly sampling was higher than most fixed-period (monthly and 6-week) plus event sampling strategies. For both sets of constituents, additional samples collected by storm chasing or single-stage samplers had relatively similar costs; storm chasing was slightly more expensive than single-stage samples for the limited set of constituents, but the opposite was true for the extensive set. The relatively similar costs were a result of more samples being collected with single-stage samplers (~9 additional samples per year) than with storm chasing (~6 additional samples per year).

For the limited set of constituents, the primary cost of the strategy was personnel time to prepare, collect, and handle the samples. Personnel costs would become much more important if the hourly rate was significantly increased from the $15 per hour. For the extensive set constituents, analytical costs dominated all other costs.

5. DISCUSSION

5.1 "True" Water Quality Versus Actual Water Quality

Actual loads and maximum concentrations in any stream cannot be determined exactly without accurate continuously recorded flow and concentration data. The data used to estimate the "true" loads (and yields) and maximum concentrations in this study were collected as continuously as economically possible: fixed-period samples were collected throughout the entire period (at least semimonthly during the open-water period and monthly during winter) and throughout most high-flow events [Graczyk et al., 1993]. Therefore, use of the integration method to compute loads with these data and determination of maximum concentrations from the extensive data sets are thought to provide the best approximations of the true values possible.

5.2 Most Effective Sampling Strategy

The results shown in Figs. 2 and 3 and summarized in Table 3 can be used to suggest the magnitude of errors expected in estimating annual yields (when a regression approach is used) and maximum concentrations in small streams (less than ~100 km^2) and the most effective sampling strategy. Regardless of the sampling strategy used, use of the regression approach to estimate annual loads for small streams with a small number of samples per year (~30 or less) is inherently imprecise and can result in significant biases in annual load estimates. The smallest errors one can expect with either 1 or 2 years of data are ~30 to 40% (median absolute errors, and standard deviation of the errors). With more than 2 years of data, the smallest errors one can expect remains about 30% (median absolute errors) with a standard deviation in the errors of about 45% (square root of the variance). The magnitudes of the errors found in this study are larger than those found using the regression approach for large rivers [Dolan et al. 1981; Preston et al. 1989], but of similar magnitude to those found for small streams using either hourly or daily average streamflow in the regression [Walling and Webb 1981,1988]. However, Walling and Webb found consistent negative biases in the estimated annual loads compared to the positive biases found in this study. Maximum concentrations were consistently underpredicted with all of the strategies examined, especially for the fixed-period strategies that were negatively biased 50 to 80%. The negative bias was minimized if samples were collected during increasing or near peak flow.

So, given a limited budget, what is the most cost-effective sampling strategy to estimate maximum concentrations and annual loads in small, flashy streams? On the basis of the results from this study, the answer depends on whether you examine maximum concentrations or annual loads and depends on the duration of the study. If you are most interested in maximum concentrations, any strategy that includes samples at or prior to peak flow regardless of the study duration is best. Therefore, a minimal fixed-period sampling supplemented with storm-chasing or single-stage samples would be best.

If you are most interested in computing annual loads, the most cost-effective strategy depends on the duration of the study. For studies more than 2 years, fixed-period, semimonthly sampling typically produces the smallest errors (least biased and most precise). It is significantly more expensive than monthly sampling, but similar in cost to the other strategies with supplemental high-flow samples. Therefore semimonthly sampling would be the most cost-effective sampling strategy. With studies longer than 3 years, it is expected that fixed-period monthly and semimonthly sampling should result in similar errors, so, fixed-period monthly sampling may be the most cost-effective sampling strategy; however, longer in-depth studies are needed to compare these strategies. Therefore, for long-term monitoring studies that do not require load estimates in the first few years of the study, fixed-period sampling would be an appropriate sampling strategy to estimate annual loads that are unbiased and as precise as those estimated with additional high-flow samples if a regression approach using daily average flows is going to be used to estimate loads.

If the duration of the study is 2 years or less, determining which sampling strategy is most cost effective is more complicated: one must weigh the tradeoffs between biased estimates and imprecise estimates. All three fixed-period sampling strategies provided relatively unbiased annual load estimates with a regression approach; however, the variance in the errors in loads with these strategies was often quite large and much larger than the typical interannual variability in the annual loads. Fixed-period monthly and semimonthly sampling plus storm chasing, on the other hand, provided relatively precise estimates for all study durations, but the loads were overestimated by ~30 to 50%. Therefore, because interannual variability in annual loads is generally much greater than 30 to 50%, and because semimonthly plus storm chasing was only slightly better than monthly plus storm chasing and yet much more expensive, the most cost-effective sampling strategy for 1-year studies appears to be fixed-period monthly sampling supplemented with storm chasing. For 2-year studies, semimonthly sampling and monthly sampling plus storm chasing produced similar overall errors (mean square errors). Therefore, because semimonthly sampling (without additional high-flow samples) produced similarly precise estimates without consistently overestimating the annual loads, it is the most cost-effective strategy for 2-year studies.

Given the biases and imprecision in the load estimates from the regression approach, one may be inclined to use another method to estimate loads in small streams. However, because concentrations of TP and other sediment-derived constituents change quickly and because the number of samples usually collected or able to be collected is small, the regression approach may be the only viable approach.

5.3 Effects of Different High-flow Sampling Strategies

Additional high-flow samples are generally collected to detect the maximum concentrations in streams that typically occur during high flow and were shown to be needed to accurately detect maximum concentrations. In theory, one would hypothesize that additional high-flow samples, regardless of how or when they were collected, would be better than having no additional samples when using a regression approach to estimate loads using daily average streamflow. However, this study indicates these additional samples can result in not only positive biases in the annual loads but also less precise estimates. The primary reason for this is that the measured concentrations are generally higher than the actual average daily concentrations during the high-flow event. Therefore, the measured concentrations typically do not represent the day on the whole, as well as random samples collected using a fixed-sampling strategy. Therefore, when using the "big river" approach (regression using daily average streamflows), samples should be randomly collected during high-flow days rather than during parts of the events. The most effective strategy examined here to collect high-flow samples is that for storm chasing by sampling crews. Because of the inability of sampling crews to immediately respond to high-flow events, samples are less consistently collected before or near peak flow, which results in better load estimates for these types of streams.

5.4. Alternative Long-term Sampling Strategy for Small Streams

The strategies that improve the estimation of maximum concentrations decrease the accuracy and precision in annual loads estimated using a regression approach. One approach to maximize the accuracy in both of these estimations would be to modify the strategy as a study progresses. This new strategy would consist of using fixed-period monthly sampling supplemented with storm chasing for the first few (2 to 3) years (NAWQA approach), which should produce relatively accurate loads (although biased high) and only moderately (negatively) biased maximum concentrations. For the next few (3) years, fixed-period monthly sampling could be supplemented with high-flow samples collected with single-stage samplers on the rising limbs of events instead of with storm chasing, which would improve the accuracy of the estimated maximum concentrations. Then if all of the high-flow samples (storm-chasing and single-stage samples) were omitted from the regression analyses, relatively accurate and precise annual loads could be estimated for all years. In the following years, the sampling could be reduced to fixed-period monthly sampling to continue to compute annual loads while assembling a database to detect possible long-term changes in water quality.

6. SUMMARY AND CONCLUSIONS

Different temporal sampling strategies strongly affect the estimated annual loads and maximum concentrations in small streams. The most cost-effective sampling strategy to estimate maximum concentrations and annual loads in small, flashy streams depends on whether you examine maximum concentrations or annual loads and on the duration of the study. If you are most interested in maximum concentrations, any strategy that includes samples at or prior to peak flow, regardless of the study duration, is best. Therefore, a minimal fixed-period sampling supplemented with single-stage or peak-flow samples would be best. The most effective sampling strategy to estimate loads in small streams depends on the duration of the study. For 1-year studies, fixed-period monthly sampling supplemented with storm chasing was most cost-effective because it results in the most precise annual loads, even though this approach usually overestimated annual loads by 25 to 50%. For studies of 2 to 3 years, fixed-period semi-monthly sampling provided not only the least biased but also the most precise estimates.

Additional high-flow samples are commonly collected to detect maximum concentrations and help define the relation between high flow and high loads for the regression approach to estimate loading. These samples are needed to detect maximum concentrations; however, they result in not only positive biases in load estimates but also in less precise estimates because concentrations in these samples do not represent average daily concentrations, especially if they are consistently collected prior to or near peak flow. The most effective strategy for collecting high-flow samples that represent a daily average concentration was the approach described for storm chasing because sampling crews generally do not respond quickly enough to consistently bias the samples toward the high concentrations during increasing flow.

Acknowledgements. We would like to thank R. Bannerman, Wisconsin Department of Natural Resources and the Environmental Studies Section and the National Water-Quality Assessment Program, USGS, for their funding, data, and assistance with the data used in this study and E. Roerish for assistance in data analysis.

7. REFERENCES

Cohn, T. A., Recent advances in statistical methods for the estimation of sediment and nutrient transport in rivers, Reviews of Geophysics, Supplement, U.S. National Report to International Union of Geodesy and Geophysics 1991-1994, 1117-1123, 1995.

Cohn, T. A., L. L. DeLong, E. J. Gilroy, R. M. Hirsch, and D. K. Wells, Estimating constituent loads, Water Resour. Res., 25(5), 937-942, 1989.

Dolan, D. M., A. K. Yui and R. D. Geist, Evaluation of river load estimation methods for total phosphorus, J. Great Lakes Res., 7, 207-214, 1981.

Edwards, T. K., and G. D. Glysson, Field methods for measurement of fluvial sediment: U.S. Geol. Surv. Open-File Report 86-531, 118 p., 1988.

Gilliom, R. J., W. M. Alley, and M. E. Gurtz, 1995, Design of the National Water-Quality Assesment program: Occurrence and distribution of water-quality conditions, U.S. Geol. Surv. Circ. 1112, 1995.

Graczyk, D. J., J. F. Walker, S. R. Greb, S. R. Corsi, and D. W. Owens, Evaluation of nonpoint-source contamination, Wisc.: Selected data for 1992 water year, U.S. Geol. Surv. Open-File Report 93-630, 1993.

Guy, H. P., and V. W. Norman, Field methods for measurement of fluvial sediment, U.S. Geol. Surv. Techniques of Water-Res. Invest., book 3, chap. C2, 1970.

Hirsch, R. M., W. M. Alley, and W. G. Wilber, Concepts for a National Water-Quality Assessment program, U.S. Geol. Surv. Circ. 1021, 1988.

Illinois Environmental Protection Agency, Bureau of Water, Illinois water quality report, 1994 and 1995, vol. 1, IEPA/BOW/96-060a, 1996.

Porterfield, G., Computation of of fluvial-sediment discharge, U.S. Geol. Surv. Techniques of Water-Res. Invest., book 3, chap. C3, 1972.

Preston, S. D., V. J. Bierman Jr., and S. E. Silliman, An evaluation of methods for the estimation of tributary mass loads, Water Resour. Res., 25(6), 1379-1389, 1989.
Robertson, D. M. and E. D. Roerish, Influence of various water quality sampling strategies on load estimates for small streams, Water Resour. Res., 35(12), 3747-3759, 1999.
Tiegs, C., State of Wisconsin surface water quality monitoring data, 1986, WR222-90, Wisconsin Department of Natural Resources, 1986.
U.S. Geological Survey, National Water Information System (NWIS), U.S. Geol. Surv. Fact Sheet FS-027-98, 1998.
Walling, D. E., and B. W. Webb, The reliability of suspended sediment load data. In Erosion and Sediment Transport Measurement. Proc. Florence Symp., June 1981, IAHS Publ. no. 133, 177-194, 1981.
Walling, D. E., and B. W. Webb, The reliability of rating curve estimates of suspended sediment yield: some further comments. In Sediment Budgets. Proc. Porto Alegre Symp., Dec. 1988, IAHS Publ. no. 174, 337-350, 1988.
Walker, W. W., Simplified procedures for eutrophication assessment and prediction: User manual, Instruction Report W-96-2, U.S. Army Corps of Engineers, 1996.
Wisconsin State Laboratory of Hygiene, Manual of analytical methods, inorganic chemistry unit, Environmental Sciences Section, University of Wisconsin-Madison, 1993.