JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION

Vol. 48, No. 6 AMERICAN WATER RESOURCES ASSOCIATION December 2012

INFLUENCE OF SAMPLING FREQUENCY ON ESTIMATION OF ANNUAL TOTAL PHOSPHORUS AND TOTAL SUSPENDED SOLIDS LOADS[1]

Amber Spackman Jones, Jeffery S. Horsburgh, Nancy O. Mesner, Ronald J. Ryel, and David K. Stevens[2]

ABSTRACT: The determination of sediment and nutrient loads is typically based on the collection and analysis of grab samples. The frequency and regularity of traditional sampling may not provide representation of constituent loading, particularly in systems with flashy hydrology. At two sites in the Little Bear River, Utah, continuous, high-frequency turbidity was used with surrogate relationships to generate estimates of total phosphorus and total suspended solids concentrations, which were paired with discharge to estimate annual loads. The high frequency records were randomly subsampled to represent hourly, daily, weekly, and monthly sampling frequencies and to examine the effects of timing, and resulting annual load estimates were compared to the reference loads. Higher frequency sampling resulted in load estimates that better approximated the reference loads. The degree of bias was greater at the more hydrologically responsive site in the upper watershed, which required a higher sampling frequency than the lower watershed site to achieve the same level of accuracy in estimating the reference load. The hour of day and day of week of sampling impacted load estimation, depending on site and hydrologic conditions. The effects of sampling frequency on the determination of compliance with a water quality criterion were also examined. These techniques can be helpful in determining necessary sampling frequency to meet the objectives of a water quality monitoring program.

(KEY TERMS: turbidity; nutrients, monitoring; sampling frequency; total phosphorus; suspended sediment; load estimation.)

Jones, Amber Spackman, Jeffery S. Horsburgh, Nancy O. Mesner, Ronald J. Ryel, and David K. Stevens, 2012. Influence of Sampling Frequency on Estimation of Annual Total Phosphorus and Total Suspended Solids Loads. *Journal of the American Water Resources Association* (JAWRA) 48(6): 1258-1275. DOI: 10.1111/j.1752-1688.2012.00684.x

INTRODUCTION

Water quality constituent loadings are generally determined through the collection and analysis of concentration grab samples along with instantaneous estimates of discharge. For most water quality monitoring programs, the frequency of grab sampling requires a balance between the necessary resolution to estimate accurate loads and the resource costs of sampling (Kronvang and Bruhn, 1996; Coynel *et al.*, 2004). Miller *et al.* (2007) and Brauer *et al.* (2009) observe that monitoring programs often have arbitrary sampling frequencies that may not match the

---

[1] Paper No. JAWRA-12-0058-P of the *Journal of the American Water Resources Association* (JAWRA). Received March 2, 2012; accepted July 17, 2012. © 2012 American Water Resources Association. **Discussions are open until six months from print publication.**

[2] Respectively, Research Engineer (Jones), Research Assistant Professor (Horsburgh), Professor (Stevens), Utah Water Research Laboratory, Utah State University, 8200 Old Main Hill, Logan, Utah 84322-8200; and Associate Dean (Mesner), College of Natural Resources, Associate Professor (Ryel), Department of Wildland Resources, College of Natural Resources, Utah State University, Logan, Utah 84322-5200 (E-mail/Jones: amber.jones@usu.edu).

temporal variability of the constituent of concern. Furthermore, the frequency required for grab sampling to be representative of constituent behavior may be logistically infeasible due to the number of samples that will have to be collected and analyzed (Coynel *et al.*, 2004). Water quality models require concentration observations for calibration and also suffer from sparse concentration datasets (Neilson and Chapra, 2003). Compliance with water quality regulations is often determined using concentrations of grab samples and resulting loads, even when important periods for constituent transport may be omitted (Jordan *et al.*, 2007). Although many aspects of water quality monitoring and modeling have improved, sampling frequency is now, and is likely to remain, a limiting factor in load estimation and water quality modeling (de Vries and Klavers, 1994; Kirchner *et al.*, 2004; Johnes, 2007).

If sensor technology is available, or if a surrogate relationship can be developed between a sensor measurement and a constituent of interest, high frequency monitoring using *in situ* sensors can enhance traditional grab sampling. High frequency monitoring captures periods that are often overlooked by routine sampling and overcomes the logistic challenges required for representative sampling (Grayson *et al.*, 1997; Christensen *et al.*, 2002; Tomlinson and De Carlo, 2003; Kirchner *et al.*, 2004; Jordan *et al.*, 2007). Several authors recommend high frequency, continuous monitoring to reduce uncertainty in load calculation resulting from infrequent sampling and biased estimation methods (Walling and Webb, 1981; Ferguson, 1987; de Vries and Klavers, 1994; Quilbe *et al.*, 2006; Johnes, 2007; Brauer *et al.*, 2009). In addition, high-frequency monitoring eliminates the need to select one of many complex equations for calculation of loads, most of which are designed to overcome sparse data. For example, Richards and Holloway (1987), Kronvang and Bruhn (1996), Phillips *et al.* (1999), and Coynel *et al.* (2004) looked at multiple load calculation equations, and all affirmed that, regardless of calculation method, results improve as sampling frequency increases.

An appropriate sampling regime depends on constituent behavior, watershed characteristics, and study objectives (Miller *et al.*, 2007; Brauer *et al.*, 2009) and should be based on a scientific rationale such as an acceptable level of uncertainty in load calculation for the constituent of concern (Birgand *et al.*, 2010a). As various studies have generated recommendations specific to constituents, watersheds, and study objectives (e.g., Schleppi *et al.*, 2006; Stelzer and Likens, 2006; Birgand *et al.*, 2010a; Cassidy and Jordan, 2011), it can be difficult to determine an acceptable sampling frequency *a priori*. Sampling frequency may be an especially important factor in load calculation for small watersheds (Brauer *et al.*, 2009; Duvert *et al.*, 2011), flashier flow regimes (Whyte and Kirchner, 2000; Stelzer and Likens, 2006), and readily soluble or sediment-associated constituents (Miller *et al.*, 2007). Schleppi *et al.* (2006) note that there is "…no definitive agreement on which sampling and integration methods are the best for large rivers and the case of small and responsive streams is even more open."

This article examines the effect of sampling frequency on annual load calculations using random subsets of high-frequency concentration estimates to simulate periodic grab sampling at different frequencies. The timing of sampling, which few studies have addressed, is investigated. We also examine the effects of sampling frequency on the determination of compliance with a regulatory criterion, an application that has not been a focus of previous analyses of sampling regimes. In related research, surrogate regression relationships were developed for two sites in the Little Bear River in northern Utah using turbidity as an explanatory variable for total phosphorus (TP) and total suspended solids (TSS) that consider censored data (i.e., values below the detection limit) as well as hydrologic conditions (Jones *et al.*, 2011). These relationships were used to construct continuous, high-frequency (half hour interval) time series of estimated TP and TSS concentrations at two sites. In this article, we describe the results of decimating these synthetic concentration records at varying time intervals to create time series subsets that simulate periodic grab sampling. Each subset was used to calculate an associated annual load for two years of data. These loads were then compared to the reference loads calculated using the entire synthetic concentration record. We begin by describing the study area and the sites at which loads were calculated. We then detail the methods that were used in deriving the concentration time series, decimating the datasets, calculating loads, and evaluating the results. In the Results and Discussion section, we relate the results of the load calculations for each scenario, compare them to the reference loads, compare results between the two sites, and examine the results in the context of a regulatory nutrient criterion.

STUDY AREA

The Little Bear River watershed (740 km^2, Figure 1) is located in northern Utah and is a major tributary of the Bear River that flows into the Great Salt Lake. Elevations in the watershed range from 1,340 to 2,700 m. The headwaters of the watershed

FIGURE 1. Little Bear River Watershed.

are primarily forest and range land. As the river flows northward through the towns of Avon, Paradise, and Hyrum, most of the adjacent land is agricultural including crops and livestock grazing. There are two impoundments within the watershed, Porcupine Reservoir and Hyrum Reservoir, which are operated to supply summer irrigation water. Below Hyrum dam, the river flows northwest through lower gradient agricultural land and the towns of Wellsville and Mendon before draining into an arm of Cutler Reservoir.

The average annual precipitation in the lower watershed is around 450 mm (recorded at the Logan 5 SW Experimental Farm Weather Station, Utah Climate Center, http://climate.usurf.usu.edu, accessed 2009) and is approximately 1,000 mm in the upper watershed, as recorded at the Little Bear River Snowpack Telemetry (SNOTEL) site (U.S. Department of Agriculture, National Resource Conservation Service, http://www.wcc.nrcs.usda.gov/nwcc/site?sitenum=582&state=ut, accessed 2009). Most of the precipitation occurs as snowfall, and the flow regime in the watershed is driven by spring snowmelt, with hydrograph peaks occurring in late spring.

This article examines loads calculated at two locations in the Little Bear River, indicated in Figure 1. The first site is the Little Bear River at Paradise, located in the upper watershed above Hyrum Reservoir. The second site is the Little Bear River at Mendon, located in the lower watershed near the river's terminus. The two sites were selected for their differing characteristics. Above Paradise, there are agricultural diversions, and the river passes through some agricultural land, but relative to Mendon, the river is less regulated, higher gradient, and less impacted by human activity. In contrast, above Mendon, the river is controlled by reservoir releases and influenced by agricultural return flows, discharge from wastewater treatment lagoons, and an increasingly agriculturally developed landscape. Approximately, 4% of the land above Paradise is agricultural, whereas between Paradise and Mendon, agriculture accounts for about 50% of total land use. Soil characteristics and instream sediment dynamics also differ between the two sites. Mendon is located in a lacustrine valley with finer soils that remain in suspension, whereas the suspended sediment at Paradise is coarser and more likely to settle (Soil Survey Staff NRCS USDA, 2008). Differences between the two sites are also evident in discharge records. Mendon generally has greater base flow discharge with attenuated peaks, whereas the discharge at Paradise is flashier. For the two years that comprise the period of this study, the mean discharge at Paradise was 2.5 cms with a maximum of 29 cms (recorded at U.S. Geological Survey [USGS] gage 10105900 Little Bear River at Paradise, Utah, http://waterdata.usgs.gov/nwis/dv?referred_module=sw&site_no=10105900, accessed 2009); and at Mendon, the average discharge was 3.5 cms with a maximum of 11 cms (determined from data collected at our monitoring site).

METHODS

Synthetic Discharge and Concentration Time Series

For both sites, high-frequency discharge values were matched in time with TSS and TP concentrations estimated from turbidity to calculate annual loads. At Paradise, instantaneous discharge is measured at 15-min increments by the USGS (gage 10105900). Gaps in the data were filled by interpolation (if the period of the gap was less than 48 h) or by substituting the daily average discharge (as obtained from the USGS daily record) for all values on that day. As the concentration time series consists of values every 30 min, only the discharge observations matching concentration values in time were used to calculate loads. The time series of discharge at Paradise is shown in Figure 2.

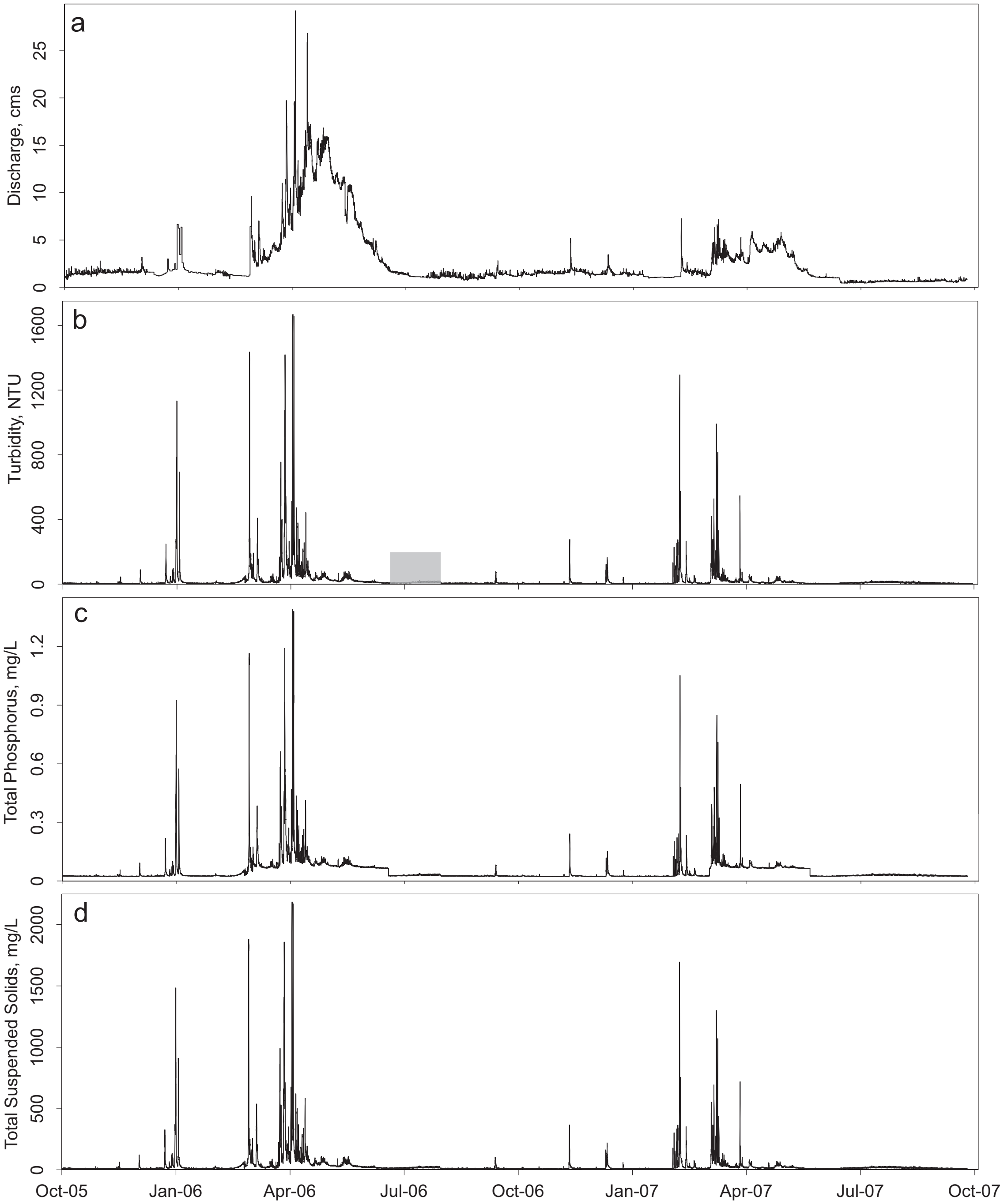

FIGURE 2. Half-Hourly Discharge (a), Turbidity (b), Estimated TP Concentration (c), and Estimated TSS Concentration (d) at Paradise. Period of substituted data is shaded in panel (b).

At Mendon, water level was measured every half hour using a KWK Technologies (Spokane, WA) SPXD-600 SDI-12 Pressure Transducer. The water level measurements were paired with field observations of water surface elevation and discharge to develop a stage-discharge relationship. Field discharge measurements were made using either a Teledyne RD Instruments (Poway, CA) StreamPro Acoustic Doppler Current Profiler or a velocity meter attached to a wading rod or a bridge cart. To ensure that the stage-discharge relationship ($R^2 = 0.98$) was representative, discharge measurements were made throughout the course of the study ($n = 7$). All estimated flows during the period of study fell within the range of the manually measured discharge values. The stage-discharge relationship was then used to calculate a half-hourly time series of discharge estimates. There were a few short periods of missing data at this site as well, and values were estimated using linear interpolation. Periods requiring interpolation did not coincide with precipitation events. The time series of discharge at Mendon is shown in Figure 3.

Concentrations of TP and TSS were estimated using site specific relationships with turbidity as detailed in Jones *et al.* (2011) (correlations shown in Figure 4). In short, least squares regression was used to develop turbidity-TP and turbidity-TSS relationships at both sites, incorporating censored data using maximum likelihood estimation techniques. In addition to turbidity, other explanatory variables (e.g., discharge, temperature, hour of day) were examined for significance. Variables representing hydrologic conditions (spring snowmelt runoff or a storm event) were also investigated. The selected model equations (Table 1) provided the minimum root mean square error values and included explanatory variables that had p-values within the 95% significance level. Although the results in Figure 4 and Table 1 were published in Jones *et al.* (2011), we include them here to show the strength of the relationships used to determine concentration estimates. The correlations in Figure 4 do not include regression lines because of the complexity introduced by categorical variables in the regressions.

Turbidity was measured every half hour at each site using a Forest Technology Systems (Victoria, BC, Canada) DTS-12 SDI-12 Turbidity Sensor. A small number of gaps in the turbidity data having lengths less than four hours were filled by linear interpolation. At Paradise, probe malfunction resulted in a period of approximately six weeks of missing turbidity data during the summer of 2006 (Figure 2). As turbidity is low and relatively constant during the summer at that site (as confirmed by subsequent years of turbidity data), the gap was filled with data from the same dates of an adjacent year. Turbidity at Paradise was not affected by snowmelt during this period, and no significant precipitation was observed. The turbidity data series were used as input to the equations in Table 1 to generate high-frequency time series of concentration estimates. Time series of turbidity and estimated concentrations of TP and TSS at Paradise are shown in Figure 2 and at Mendon in Figure 3. Steps in the estimated TP time series are a result of the regression equations varying between periods of base flow and spring snowmelt runoff.

Calculation of Reference Loads

Several studies have assembled reference datasets that were then artificially decimated to compare the effect of sampling frequency on constituent load estimates. Reference datasets for TP and TSS in some studies were based on infrequently sampled data (i.e., at a daily frequency) (de Vries and Klavers, 1994; Robertson and Roerish, 1999; Robertson, 2003; Johnes, 2007). Others generated higher-frequency data through interpolation (Kronvang and Bruhn, 1996; Granger *et al.*, 2009), sample collection using an auto-analyzer (Cassidy and Jordan, 2011), or using discharge rating curves (Webb *et al.*, 2000), which have been shown to be unsatisfactory estimators of TP and TSS (Phillips *et al.*, 1999; Robertson and Roerish, 1999; Quilbe *et al.*, 2006; Johnes, 2007; Jordan *et al.*, 2007). In this study, we used methods similar to those of Walling and Webb (1981), Phillips *et al.* (1999), and Duvert *et al.* (2011) to create high-frequency estimates of TSS and TP concentrations calculated from turbidity.

Using the complete sets of discharge and concentration data (half hourly), annual loads were determined for each water year at each site. These loads provided a reference to which loads estimated from scenario subsamples could be compared. Reference loads were calculated according to Equation (1):

$$W = \sum_{i=0}^{n} Q_i C_i x, \tag{1}$$

where W is the total annual load (kg), Q_i represents the incremental discharge (cms), C_i represents the incremental concentration of TP or TSS (mg/L), x is a factor to convert to kg per appropriate time period, and n is the total number of paired discharge and concentration estimates in one year. Equation (1) is a simple method for determining constituent loads as used by Coynel *et al.* (2004), Schleppi *et al.* (2006), Granger *et al.* (2009), and Duvert *et al.* (2011).

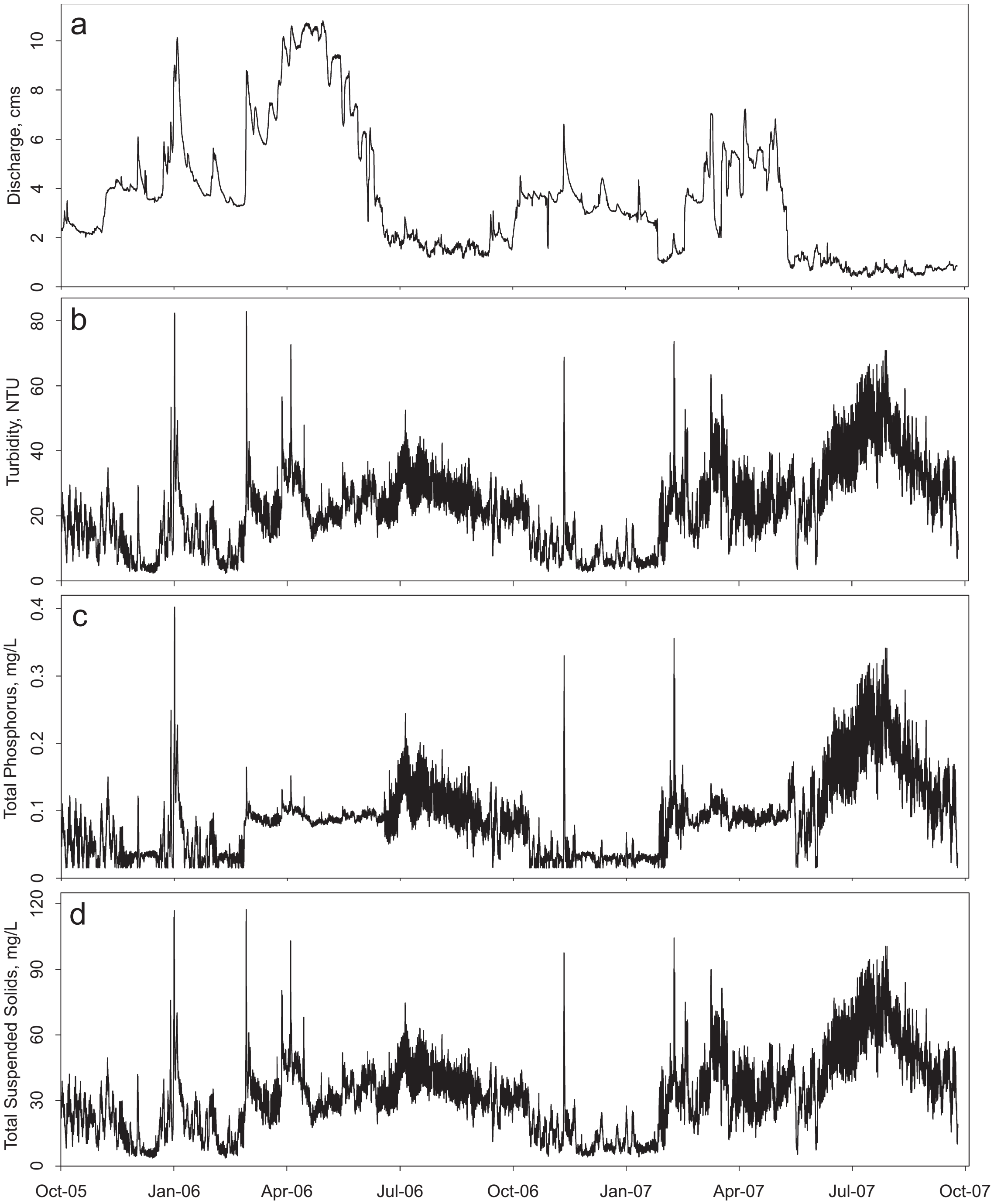

FIGURE 3. Half-Hourly Discharge (a), Turbidity (b), Estimated TP Concentration (c), and Estimated TSS Concentration (d) at Mendon.

Scenario Generation

To test how much information is lost by sampling at different temporal frequencies, several scenarios were identified. There are many scenarios that could be generated based on sampling methods currently used in water quality monitoring programs. For example, a number of studies have used various strategies for targeting sampling to events such as storms (Richards and Holloway, 1987; Kronvang and Bruhn, 1996; Robertson and Roerish, 1999; Johnes, 2007; Miller *et al.*, 2007). These approaches focus sampling on periods that would not be captured with fixed period sampling, but there is little agreement about which approaches are most appropriate (Richards and Holloway, 1987; Schleppi *et al.*, 2006).

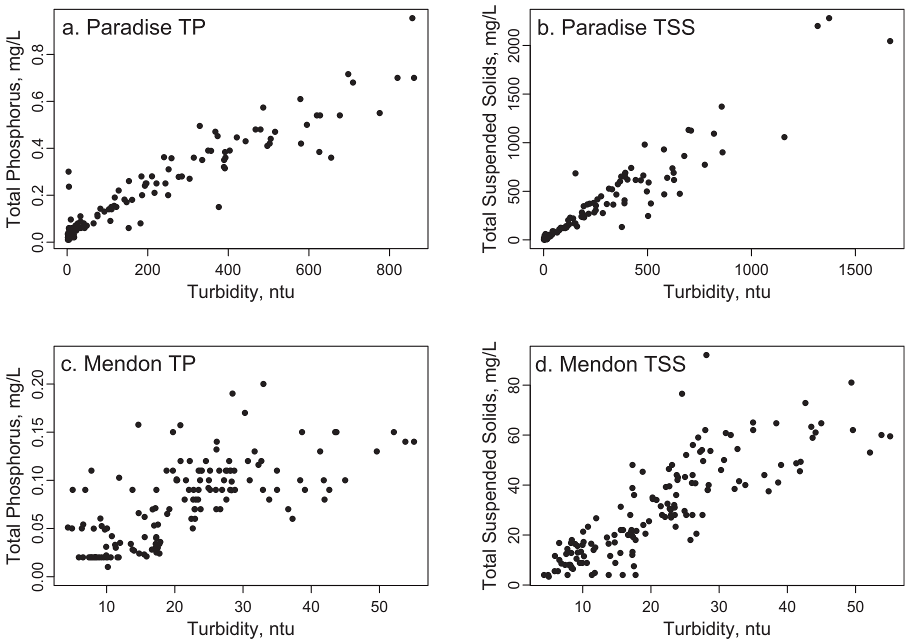

FIGURE 4. Scatter Plots Showing Correlation Between Turbidity and TP and Turbidity and TSS at Paradise and Mendon. Figure reproduced from Jones *et al.* (2011).

TABLE 1. Relationships Used to Derive Continuous Concentration Time Series (reproduced from Jones *et al.,* (2011)).

Site	Constituent	Equation	RMSE	*p*-value	Standard Error
Paradise	Total phosphorus	$TP = 0.0209 + 0.000798 \times Turb + 0.0386 \times Z$	0.069 mg/L	$Turb$: $<10^{-6}$ Z: 8.71×10^{-4}	$Turb$: 2.67×10^{-5} Z: 1.16×10^{-2}
	Total suspended solids	$TSS = 3.58 + 1.31 \times Turb$	117 mg/L	$Turb$: $<10^{-6}$	$Turb$: 3.12×10^{-2}
Mendon	Total phosphorus	$TP = -0.0341 + 0.0053 \times Turb + 0.0949 \times Z - 0.00404 \times Turb \times Z + 0.0832 \times Y - 0.000871 \times Turb \times Y$	0.027 mg/L	$Turb$: $<10^{-6}$ Z: $<10^{-6}$ $Turb \times Z$: $<10^{-6}$ Y: 1.38×10^{-3} $Turb \times Y$: 5.24×10^{-3}	$Turb$: 4.39×10^{-4} Z: 1.13×10^{-2} $Turb \times Z$: 4.86×10^{-4} Y: 2.60×10^{-2} $Turb \times Y$: 3.12×10^{-3}
	Total suspended solids	$TSS = 0.341 + 1.41 \times Turb$	10.8 mg/L	$Turb$: $<10^{-6}$	$Turb$: 7.70×10^{-2}

Notes: Variable description: *TP*, total phosphorus, mg/L; *TSS*, total suspended solids, mg/L; *Turb*, turbidity, NTU; *Z*, categorical variable for spring snowmelt ($Z = 1$) *vs.* base flow ($Z = 0$); Y, categorical variable for $Turb$ <10 NTU ($Y = 1$) *vs.* $Turb$ >10 NTU ($Y = 0$). RMSE, root mean square error.

Jordan *et al.* (2005) note that sampling regimes focusing on high-flow events overlook cyclic patterns occurring at low flows that are not simply noise in the concentration of phosphorus. Kronvang and Bruhn (1996) found load estimates from regular sampling to be more accurate than those of stratified high-flow sampling; and Miller *et al.* (2007) observed the greatest accuracy and precision in load estimates using the highest frequency fixed period sampling. For these reasons, and because fixed period sampling is ubiquitous in water quality monitoring programs, regular intervals were used to generate scenarios for this study. This also highlighted only the effects of sampling frequency and avoided the complication of disparately sampled concentration and discharge (e.g., Ferguson, 1987; de Vries and Klavers, 1994; Kirchner *et al.*, 2004; Quilbe *et al.*, 2006).

Scenarios were generated by artificially subsampling the half-hourly reference datasets at hourly, daily, weekly, and monthly frequencies, resulting in subsets of the discharge and concentration data from which annual loads for water years 2006 and 2007 could be calculated. Equation (1) was used to estimate an annual load for all scenario subsets. The following paragraphs describe the subsets of data that were generated (also summarized in Table 2).

Calculation of Annual Loads Using Daily Subsamples. To represent sampling at a daily frequency, two types of subsets were generated. The first type was created by randomly selecting an instance of corresponding discharge and concentration within each day of the year, resulting in 365 values per year. Equation (1) was then used to calculate annual loads. To achieve a distribution of load estimates using this method, random sampling and load calculation was conducted 10,000 times. The second type of daily subset was created to examine the effects of sampling time on load estimates. To simulate consistently sampling at the same hour of the day, corresponding discharge and concentration values were selected for each hour of the day on every day of the year resulting in 24 subsets (each one containing the 365 paired discharge and concentration values for the selected hour of the day) from which annual loads were calculated.

Calculation of Annual Loads Using Weekly Subsamples. Two types of subsets were also generated to simulate weekly sampling. The first type was created by randomly selecting a single, paired discharge and concentration from within each week, resulting in a subset with one paired discharge and concentration for each week (52 paired values for each year). This was conducted 10,000 times, and 10,000 annual load estimates were calculated from the subsets. The second type of weekly sampling was designed to assess the impact of consistently sampling on a particular day of the week. For each day of the week, one pair of discharge and concentration values was randomly selected for the designated day within each week of the year (resulting in 52 paired values) from which an annual load was calculated. To obtain a distribution of results using this method, random subset selection and load calculation was conducted 10,000 times for each day of the week, resulting in 70,000 annual load estimates.

Calculation of Annual Loads Using Monthly Subsamples. Monthly sampling was simulated by randomly selecting a single, paired discharge and concentration for each calendar month, resulting in subsets containing 12 paired values of concentration and discharge for each year from which annual loads were calculated. Ten thousand monthly subsets and annual load calculations were realized.

TABLE 2. Summary of Decimated Datasets.

Subset	Frequency	Realizations
Complete	Half hourly	1
Hourly	Hourly	1
Daily by hour	Daily	24*
Randomized daily	Daily	10,000
Weekly by day	Weekly	70,000**
Randomized weekly	Weekly	10,000
Randomized monthly	Monthly	10,000

*One realization was generated for each hour of the day.

**10,000 realizations were generated for each day of the week.

RESULTS AND DISCUSSION

Frequency Comparison

To illustrate the effects of sampling frequency, Figure 5 shows series of TSS estimates derived from turbidity at Paradise during spring runoff (February-May) of 2006. The half-hourly concentration time series is shown, along with subsets of the half-hourly concentrations at decreasing sampling frequency. The hourly series consists of concentrations on the hour, whereas the daily, weekly, and monthly series are randomly selected concentrations from the half-hourly record. The hourly record shows little divergence from the half-hourly dataset. The daily concentration record appears to capture the general trend of TSS concentration, but it fails to portray the fine resolution variability. The weekly and monthly series completely miss the peaks in concentration, which are the periods of greatest contribution to total annual load. On the other hand, under a monthly or weekly sampling routine, a sample could be collected during a peak in concentration, which may lead to a significant overestimation of annual load.

Scholefield *et al.* (2005) recommend that the sampling frequency should match the scale of the processes involved. Kirchner *et al.* (2004) assert that the measurement frequency of chemical constituents should be often enough that no new information is gained by sampling more frequently. At the two sites in the Little Bear, the half-hourly concentrations do not reveal any pattern that is not observed in the hourly data, but the daily concentrations overlook behavior that is occurring within the day. For other watersheds or other constituents, making measurements more frequently than hourly or half hourly may be necessary. For example, Tomlinson and De Carlo (2003) used *in situ* measures at five-minute intervals to demonstrate the high variability in Hawaiian streams, and Granger *et al.* (2009) found that hourly concentrations of TP and TSS omitted resolution that was captured by half-hourly data.

Figures 6 and 7 summarize the results of subsampled load calculations for TP and TSS at Paradise and Mendon for water years 2006 and 2007. The categories in the plots correspond to simulated sampling

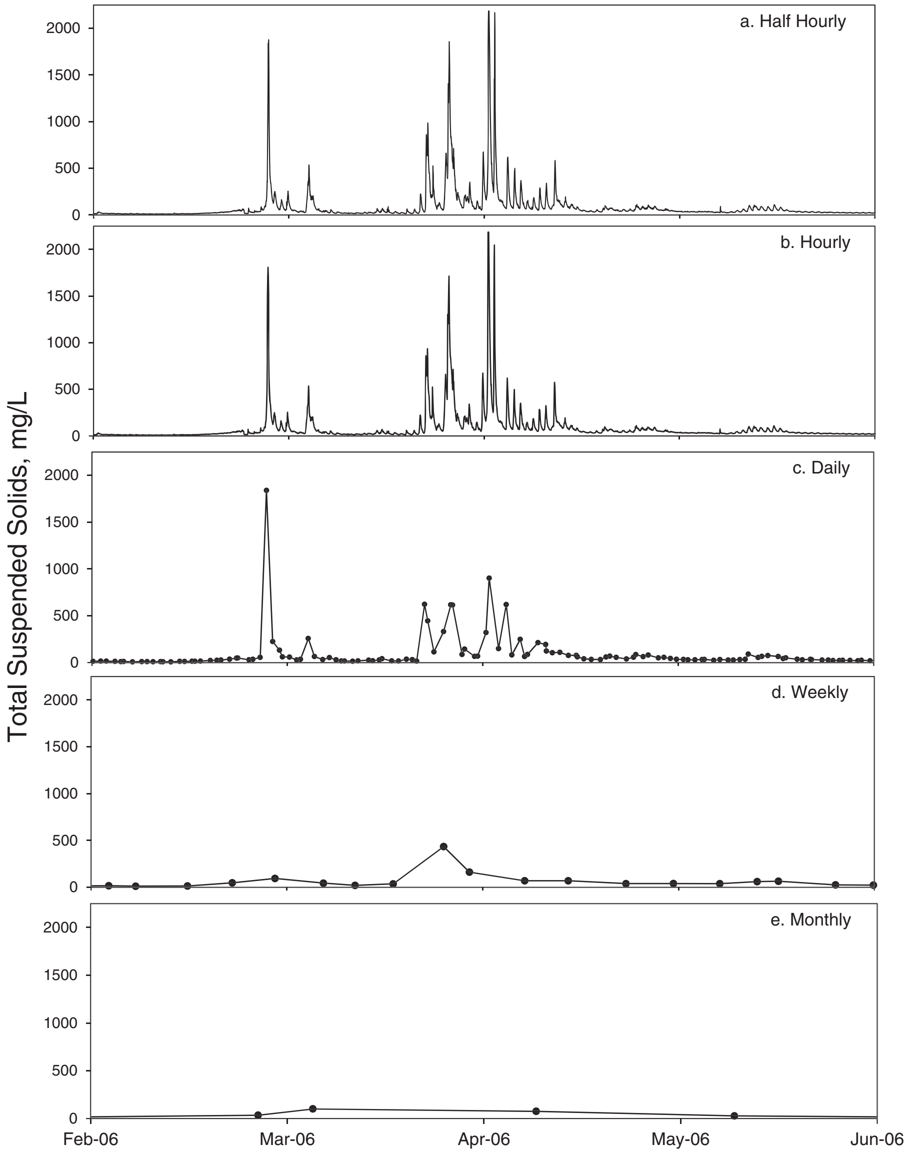

FIGURE 5. TSS Concentrations at Paradise at Varying Sampling Frequencies as Subsampled from the Half-Hourly Concentration Estimates (a). The hourly time series (b) consists of estimates made on the hour while the daily (c), weekly (d), and monthly (e) are randomly selected points.

frequency scenarios. Sampling frequency affects the range of load estimates. For all variables, sites, and years, the variability in the estimated loads increases as sampling frequency decreases because a single sample is assumed to be representative of a longer time period. The discharge and concentration at a point might not be characteristic of the surrounding period (e.g., a sample collected during a rain or snowmelt event is assumed to represent an entire month), and the resulting annual load can be skewed.

At Paradise, for both variables and years, the median loads decrease as sampling frequency decreases, indicating that less frequent sampling typically omits periods of significant constituent loading and thus underestimates annual loads. Richards and Holloway (1987), Phillips *et al.* (1999), and Coynel *et al.* (2004) also found that decreasing sampling frequency resulted in underestimated loads. At Mendon, on the other hand, the median loads for all sampling frequencies are within 5% of the reference loads.

Overall, hourly sampling frequency provides very close approximations of the reference loads at both sites, so little resolution is lost by decreasing sampling frequency to hourly. At coarser sampling frequencies, the departure of estimated loads from the reference loads varies considerably between the two sites. At Mendon, using daily subsampling, even the upper and lower adjacent levels of annual load estimates are within 5% of the reference loads for both variables and both years. Conversely, at Paradise,

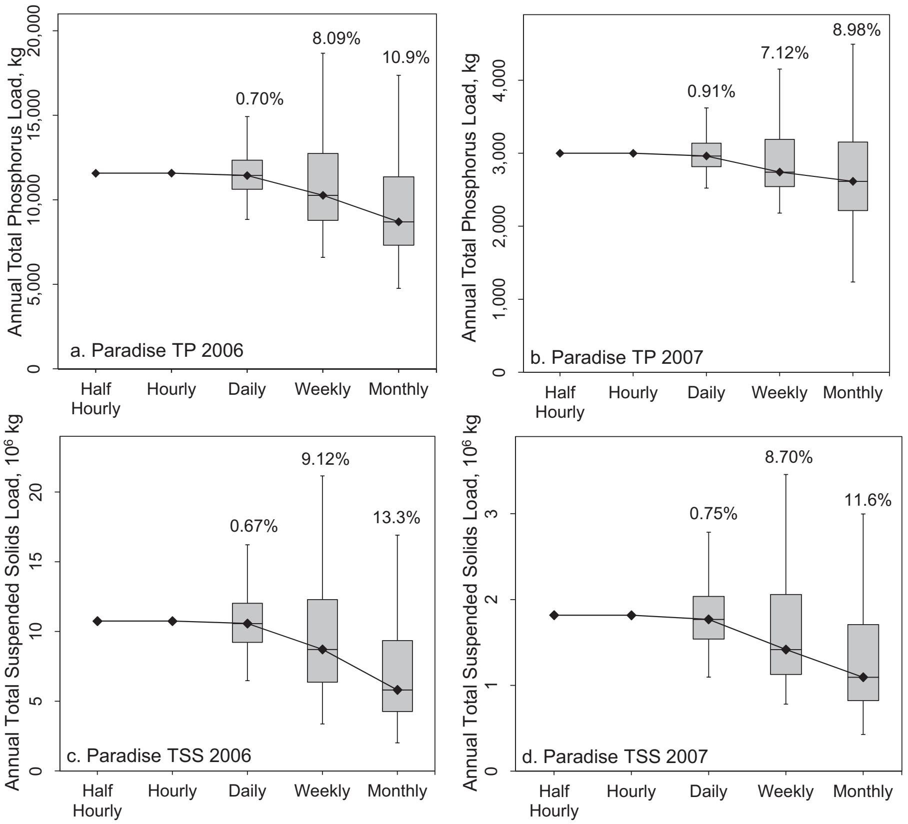

FIGURE 6. Box and Whisker Plots of the Annual Loads for Varying Sampling Frequencies at Paradise for Water Years 2006 and 2007. Half hourly represents the reference load, hourly represents the value from subsampling on the hour, and daily, hourly, and monthly represent 10,000 annual load calculations from randomly selected points within each day, week, or month. The boxes represent the first and third quartiles and the whiskers represent the lower and upper adjacent levels. The medians of each of the sets of realizations are also indicated. The percentages above the upper whisker represent the portion of calculated loads that fell above the upper adjacent level. There were no values below the lower adjacent levels.

although the medians of annual loads estimated from daily subsampling are within 5% of the reference loads, the 1st and 3rd quartiles of load estimates deviate from the reference loads by 5-15% and the upper and lower adjacent levels by 16-40%. For annual loads estimated by subsampling at a weekly frequency, at Mendon, the medians and 1st and 3rd quartiles are all within 5% of the reference loads, whereas at Paradise, the medians vary from the reference loads by 8-22%. Subsampling at a monthly frequency resulted in medians of load estimates within 5% of the reference loads at Mendon, whereas at Paradise, the medians of the load estimates from monthly subsampling underestimate the reference loads by 13-45%. In addition, at Mendon, there are fewer values falling above the upper adjacent levels than at Paradise.

At Paradise, the annual load estimates of TSS are more variable than those of TP. For weekly and monthly subsampling, the medians of TSS load estimates underestimate the reference loads by more than 20%, whereas the analogous medians of TP load estimates are within 8-25% of the reference loads. Using daily subsampling, the 1st and 3rd quartiles of TP load estimates are within 10% of the reference loads, but the quartiles for TSS load estimates are all greater than 10%. Richards and Holloway (1987) and Robertson (2003) also observed larger ranges for TSS load estimates than for TP load estimates. The difference between variables was not observed at Mendon, and no prominent difference between the two water years was observed apart from the differing scales as TP and TSS loads were greater in 2006 than in 2007. Water year 2006 was a relatively high-flow year in the Little Bear due to a considerable snowpack and favorable conditions during runoff, whereas precipitation and discharge were both low in water year 2007.

Probability of Achieving the Reference Loads

Although 10,000 load estimates were generated by randomly subsampling at daily, weekly, and monthly

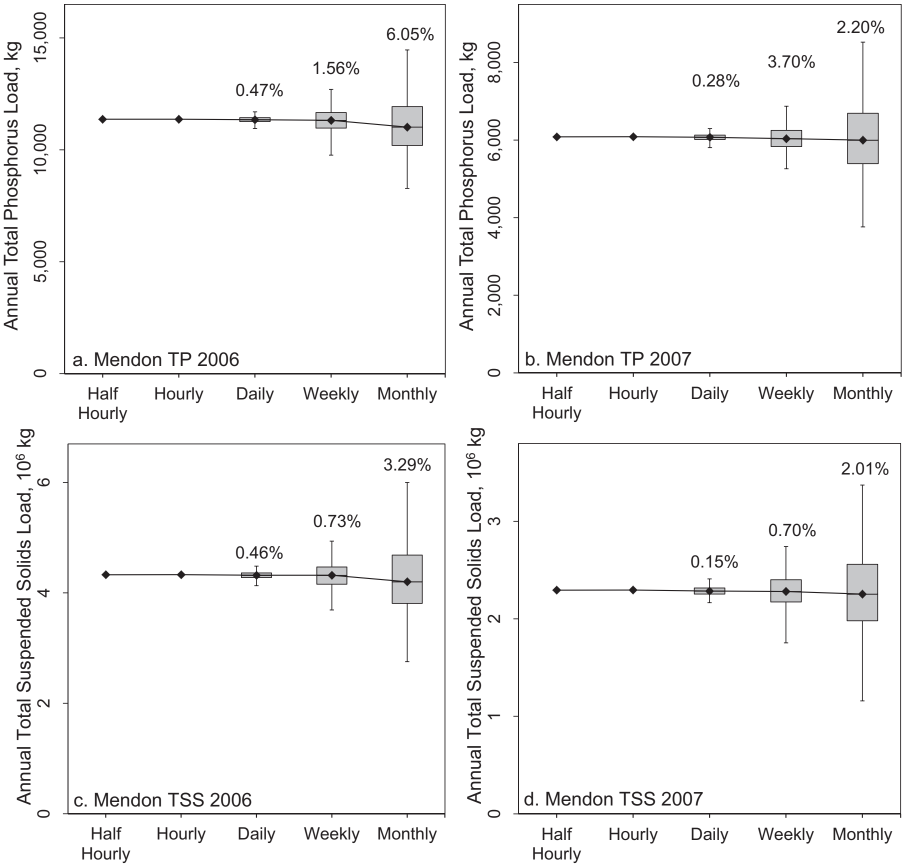

FIGURE 7. Box and Whisker Plots of the Annual Loads for Varying Sampling Frequencies at Mendon for Water Years 2006 and 2007. See Figure 6 for detailed plot description.

time scales, in reality, only one annual load estimate could be made using real sampling data, regardless of its frequency. Using the 10,000 load estimates for daily, weekly, and monthly frequencies, we examined the likelihood of a single load estimate falling within certain thresholds of the reference load. In other words, we asked how probable it is that we will be close to the simulated true loading if we sample at the given frequency. Thresholds of 5 and 50% were selected to represent being very close to the reference load and being "within the ballpark" of the reference load, respectively. Several studies used 20% or more as an acceptable error from the reference load (Robertson and Roerish, 1999; Robertson, 2003; Coynel *et al.*, 2004; Brauer *et al.*, 2009; Duvert *et al.*, 2011), but we think that more accurate loads are achievable. The results (reported in Table 3) further reveal differences between the two sampling sites. At Mendon, the probability of being within 5% of the reference load is 1.0 for sampling at a daily frequency, 0.50-0.72 for a weekly frequency, and 0.20-0.31 for monthly sampling. At Paradise, on the other hand, daily sampling only has a probability of 0.19-0.46 of achieving a load estimate within 5% of the reference load. At Mendon, it is very probable (0.98-1.0) that loads will be within 50% of the reference load, regardless of sampling frequency. In contrast, for monthly sampling at Paradise, the probability of being within 50% of the reference load is only 0.52-0.89.

Daily by Hour Loads

The variability in loads calculated by simulating consistently sampling at the same time each day is shown in Figure 8. Although the trends are distinct for each site, they are similar across variables and years. At Paradise, loads calculated from concentrations and discharges at the end of the day (hours 16:00-24:00) are higher than those calculated for hours earlier in the day, although the difference is less dramatic for both TP and TSS in water year 2007. At the most extreme, the loads vary by 50% from sampling at different times of the day. At Mendon, the highest loads are in the early hours of the morning (hours 02:00-06:00), but overall, there is less variability throughout the day than at Paradise.

TABLE 3. Probabilities of a Single Load Estimation Occurring Within a Certain Threshold of the Reference Load.

			Probability of Being Within 5% of the Reference Load			**Probability of Being Within 50% of the Reference Load**		
			Sampling Frequency					
Site	**Variable**	**Year**	**Daily**	**Weekly**	**Monthly**	**Daily**	**Weekly**	**Monthly**
Paradise	TP	2006	0.33	0.12	0.556	1.0	0.90	0.85
	TP	2007	0.46	0.15	0.14	1.0	0.95	0.89
	TSS	2006	0.19	0.071	0.022	0.99	0.72	0.43
	TSS	2007	0.19	0.057	0.044	0.99	0.82	0.52
Mendon	TP	2006	1.0	0.72	0.31	1.0	1.0	0.99
	TP	2007	1.0	0.67	0.25	1.0	1.0	0.99
	TSS	2006	1.0	0.66	0.23	1.0	1.0	0.99
	TSS	2007	1.0	0.50	0.20	1.0	1.0	0.98

Note: Probabilities are determined using the 10,000 realizations of random load calculations.

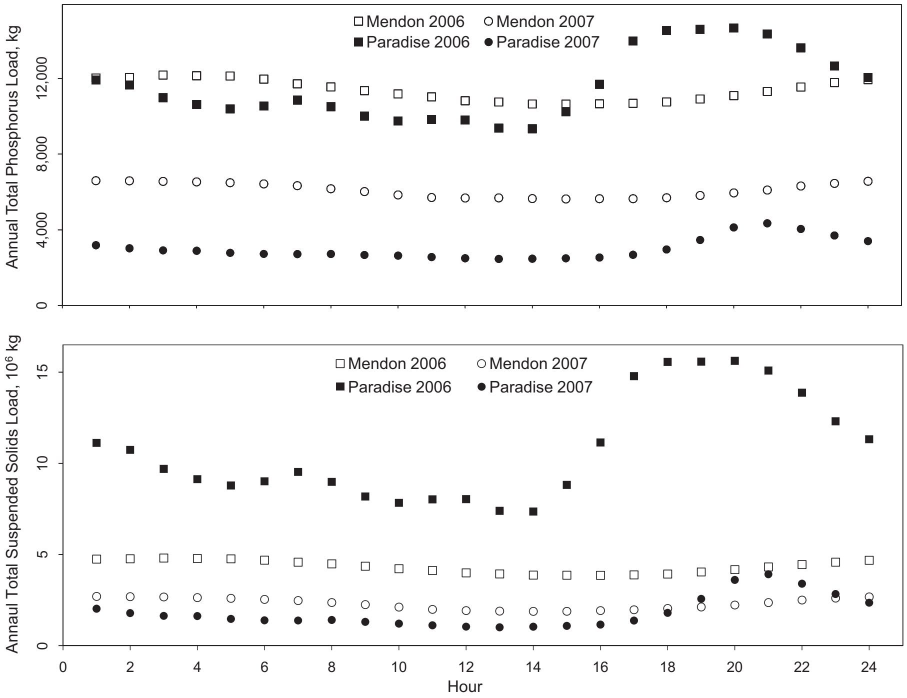

FIGURE 8. Annual Loads Calculated by Subsampling Daily at the Same Hour Each Day.

Differences in loads throughout the day are due to diurnal fluctuations in turbidity (and TP and TSS). This was confirmed by diurnal grab sampling at Paradise conducted during 2008, which revealed a broad range of TP and TSS values within a single day. During the height of spring runoff, a 24-h sampling event was conducted using an ISCO automated sampler (Teledyne Isco, Inc., Lincoln, NE), which collected one sample per hour over the 24-h period and returned TP concentrations ranging from 0.066 to 0.954 mg/L and TSS ranging from 108 to 2,450 mg/L. Samples were prepared and analyzed using the methods described in Jones *et al.* (2011). In addition, observed diurnal patterns in discharge also contributed to the differences in estimated loads.

Site-specific hydrologic conditions likely cause the differing patterns between the two sites. The timing of the response to events such as snowmelt or storms varies between the upper and lower watersheds, and the timing of reservoir releases may also affect the timing of loads at Mendon. Additional factors that could cause varying behavior within a day include changes in water temperature, evapotranspiration, and the timing of agricultural withdrawals and return flows. Scholefield *et al.* (2005) suggest that diurnal fluctuations in phosphorus concentrations may be a result of enrichment or depletion by instream biological or physical processes that respond to changes in temperature. If the diurnal variations are a result of physical processes, the intensity of variation will decrease in a downstream direction, which is the case with Paradise and Mendon. Jordan *et al.* (2007) attribute diurnal phosphorus fluctuations to rural point sources upstream of the sampling site, which may also explain some of the diurnal variability at the two locations in our study.

The diurnal turbidity patterns are similar to those observed by Loperfido *et al.* (2010), who determined that grab samples in the daytime would result in a TSS load estimate three times less than the load estimated by high-frequency sampling over a month in a small Iowa stream. Jordan *et al.* (2005) observed diurnal patterns in phosphorus at low discharge levels in a small stream in Northern Ireland that caused variability in load estimates. Duvert *et al.* (2011) found that for several small, responsive Mexican streams, most of the suspended sediment transport occurred at night and that the time of regular, daily sample collection biased load results. A similar bias is observed for both sites in the Little Bear River under the assumption that samples are typically collected during the working day (hours 09:00-05:00), which would result in an underestimation of loads for both sites.

Weekly by Day Loads

Figure 9 represents annual load estimates that simulate sampling once a week on the same day each week, but randomizing the time of day of sampling. The trends vary between sites as well as between water years. At Paradise, in water year 2006, both TP and TSS loads estimated by subsampling on Tuesdays and Wednesdays are higher and exhibit greater variability than other days of the week. In contrast, load estimates for water year 2007 at Paradise are within a smaller range and are more consistent between days of the week. At Mendon, there is no obvious pattern in loads or variability based on day of the week, water year, or variable. Although no trend is observed, the ranges of loads are still notably different between different days of the week. These results indicate that the day of the week that sampling is conducted can impact the load estimate, but

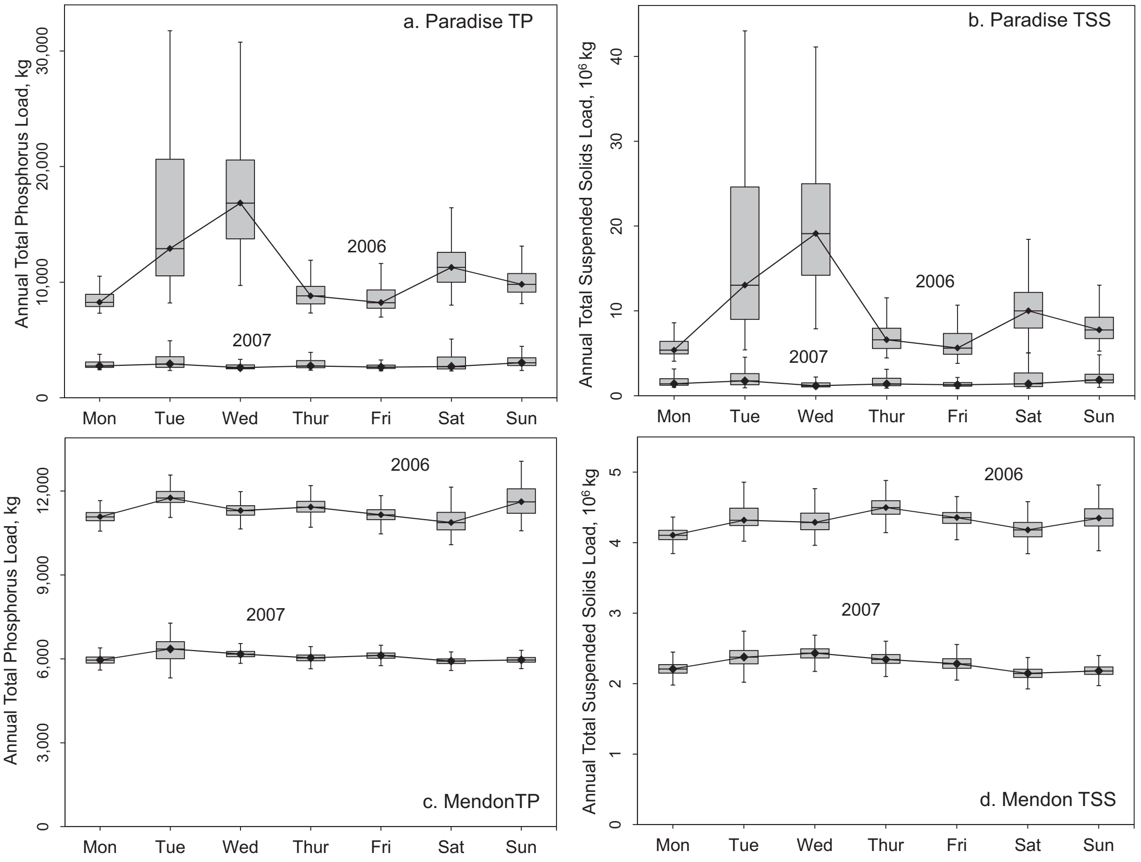

FIGURE 9. Box and Whisker Plots of Annual Loads Calculated by Consistently Subsampling on the Same Day of the Week for Water Years 2006 and 2007. Sampling frequency is weekly using a randomly selected point within a single day. Statistics are based on 10,000 realizations of annual load for each day, each site, each variable, and each water year.

we have little rationale for the observed variability. One possibility is that the differences between days of the week are related to the days on which diversions are opened or closed. Whyte and Kirchner (2000) found variability in mercury loads calculated from samples on different days of the week, which they attributed to the episodic occurrence of storms. Using precipitation data from the Logan 5 SW Experimental Farm Weather Station retrieved from the Utah Climate Center to compare days of the week, higher levels of precipitation were observed on Tuesdays and Wednesdays during the period of spring snowmelt in 2006, which may have contributed to the observed behavior at Paradise.

Site Comparison

Results highlighted in the previous sections demonstrate differences between Paradise, the upper watershed site and Mendon, the lower watershed site. At Mendon, the river generally has attenuated peaks in discharge, turbidity, and TP and TSS concentration due to reservoir operations, lower gradient, and interaction with groundwater. Paradise, on the other hand, has lower base flow and higher peaks in discharge, turbidity, and TP and TSS concentration in late spring due to the influence of snowmelt, the higher gradient of the river, and the surrounding land uses. Overall, there is more absolute variability in turbidity and resulting TP and TSS concentrations at Paradise than at Mendon, but there is greater short-term variability at Mendon, as shown in Figures 2 and 3.

In general, results show that high-frequency sampling is essential for load calculation at Paradise, but that, depending on the level of acceptable error, less frequent sampling can be conducted at Mendon. This is consistent with the findings of a number of studies comparing sampling frequency on different rivers. Smaller rivers are more responsive to precipitation and snowmelt, whereas the responses of larger rivers are more attenuated with a slower rise in discharge and higher base flow levels (Richards and Holloway, 1987). Thus, the decrease in the precision of load estimates with reduced sampling frequency is greater for smaller, more responsive rivers (Richards and Holloway, 1987; de Vries and Klavers, 1994; Kronvang and Bruhn, 1996; Phillips *et al.*, 1999; Coynel *et al.*, 2004). Duvert *et al.* (2011) report that minimum sampling frequency is highly dependent on catchment size. According to their assessment of the relationship between catchment size and sampling frequency, both Paradise (455 km^2) and Mendon (740 km^2) fall near the threshold of their recommendation for automatic *vs.* manual sampling strategies.

Also, rivers with high base flow in permeable lowlands have less variable TP and TSS, so a lower sampling frequency is acceptable compared with rivers with low base flow that transport more TP and TSS in high discharge events and require more frequent sampling (Johnes, 2007). In addition to a greater bias in load estimates at Paradise than Mendon, there was a greater degree of underestimation of loads at Paradise than at Mendon. This is consistent with studies that found loads tend to be underestimated on smaller rivers more so than larger rivers (Kronvang and Bruhn, 1996; Phillips *et al.*, 1999). Although Paradise and Mendon are located on the same river, they are separated by a reservoir, and the behavior of the river changes dramatically between the two sites. Paradise can be seen as a small river site while the attributes of a larger river could be ascribed to the Little Bear River at Mendon.

The optimal sampling frequency at a location is dependent on the responsiveness of discharge and constituent fluxes (Whyte and Kirchner, 2000; Stelzer and Likens, 2006; Duvert *et al.*, 2011), which may help to explain the differing levels of variability in load estimates between the two sites in this study. Flashiness or reactivity indices are a result of a combination of factors (e.g., slopes, soil permeability, contributing area) and can be used to quantify the level of responsiveness (Meybeck *et al.*, 2003; Robertson, 2003; Stelzer and Likens, 2006). For both Paradise and Mendon for the two years of this study, we calculated the following metrics as described by Meybeck *et al.* (2003) and Duvert *et al.* (2011): $\text{Vw}_{2\%}$, the percent of the total discharge that is transported in 2% of the time, $\text{Ms}_{2\%}$, the percent of the total material (TSS) that is transported in 2% of the time, $\text{Ts}_{50\%}$, the percentage of time in which 50% of the TSS load was discharged, and the time equivalent to that percentage. The results (Table 4) confirm our assessment of the different behavior between the two sites. Relative to Mendon, at Paradise, more of the discharge and much more of the sediment load occur during 2% of the time. At Paradise, 50% of the TSS load was discharged in 1.25% of the time, which is equivalent to approximately 9 days, whereas at Mendon, it took 114 days (15.7% of the time) to move 50% of the TSS. In comparison to results presented by Meybeck *et al.* (2003), Mano *et al.* (2009), and Duvert *et al.* (2011), for Paradise, all variables were near those of similarly sized watersheds. In contrast, at Mendon, the values for $\text{Vw}_{2\%}$ and $\text{Ms}_{2\%}$ are much lower and the value of $\text{Ts}_{50\%}$ is much greater than those of similarly sized watersheds. Relative to the watersheds assessed by Moatar and Meybeck (2007), the $\text{Vw}_{2\%}$ at Mendon was closer to those found for larger watersheds (12,000-56,000 km^2, $\text{Vw}_{2\%}$: 7.2-10.2%), whereas the $\text{Vw}_{2\%}$ for Paradise was in the range of the smaller

TABLE 4. Reactivity Index Values for the Sites in this Study and Reference Studies.

	Contributing Area, km^2	**Vw$_{2\%}$ (%)**	**Ms$_{2\%}$ (%)**	**Ts$_{50\%}$ (%)**	**Ts$_{50\%}$ Time Equivalent (days)**
Paradise	455	12.3	56.5	1.25	9.1
Mendon	740	5.93	12.1	15.7	114
Meybeck *et al.* (2003)	200-1200	11-29	28-85	0.21-6.2	-
Mano *et al.* (2009)	80-900	-	38-72	0.3-4	-
Duvert *et al.* (2011)	630	-	63	1.2	4.4

watersheds studied (1,700-1,900 km^2, Vw$_{2\%}$: 12.2-17.2). According to the classification scheme developed by Meybeck *et al.* (2003), for Ts$_{50\%}$ as an indicator of relative response time for sediment transport, Paradise lies in the very short (0.4-1.4%) range while Mendon falls in the long (8-16.5%) range. With data from additional sites, a relationship could be developed between responsiveness and desired accuracy of load estimates to determine required sampling frequency as shown by Birgand *et al.* (2010b).

Comparison with Water Quality Criteria

Although load calculations are important for estimating the overall delivery of a constituent downstream, for TP, the ultimate regulatory endpoints used in Utah waters are concentration based. Streams and rivers designated as cold-water fisheries in the state of Utah, including the Little Bear River, must meet a TP criterion of 0.05 mg/L (Utah DEQ, 2000; Utah Administrative Code R317-2, 2011). To determine the potential effects of sampling frequency on the assessment of compliance, we examined how the frequency of exceeding the criterion would vary at different sampling frequencies. For the random subsamples of concentration estimates at daily, weekly, and monthly frequencies, the frequency of exceeding the 0.05 mg/L threshold was determined for each of 10,000 realizations. The results are summarized as box plots in Figure 10.

The range of possible frequencies of exceedance broadens as sampling frequency decreases. Daily sampling returns frequencies of exceedence very similar to that of the reference dataset, and, although lower sampling frequencies are also likely to return frequencies of exceedence that are similar to the reference dataset, the range of possible results is greater. Pappas and Huang (2008) also found that frequency of exceedence of a concentration threshold for atrazine depended on the pattern and frequency of sample collection. At Paradise, for daily sampling, for the range of the upper and lower adjacent levels of 10,000 realizations, the frequency of exceedance is between 28 and 29%. For weekly sampling, the range is between 26 and 30%; and for monthly sampling, it is between 20 and 32%. At Mendon, the frequency of exceeding the criterion is higher than at Paradise. For daily sampling, the frequencies range from 69 to 72%. For weekly sampling, the range is 65-75%; and for monthly sampling, it is 55-83%. There is more variability in the range of frequencies of exceedance at Mendon than at Paradise. In the context of sampling for compliance with the concentration threshold, contrary to our findings with regard to load estimation, sampling frequency may be more important at Mendon than at Paradise. This is partly a result of the different ranges of TP levels at each site.

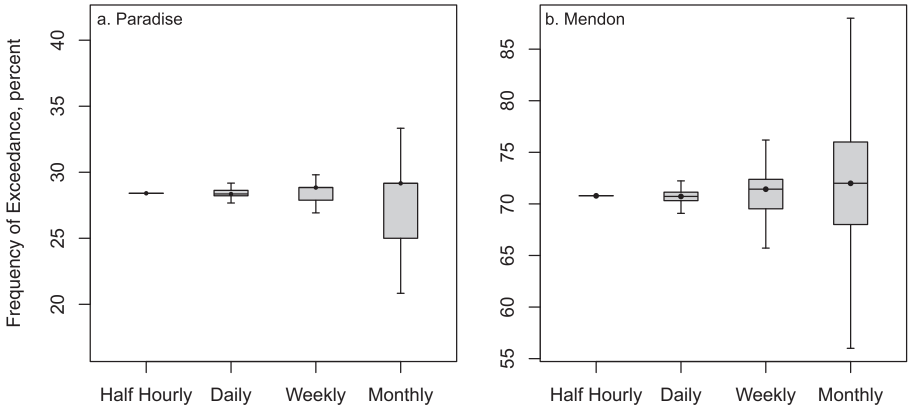

FIGURE 10. Box and Whisker Plots of the Frequency of Exceedence of Water Quality Criterion of 0.05 mg/L TP. Determined from 10,000 randomly selected concentrations for each sampling frequency.

The total range of estimated TP is greater at Paradise than at Mendon, but the median is higher at Mendon. The concentration criterion is a binary response without regard to the degree of exceedance, only whether or not exceedance occurs. Furthermore, discharge, which is integral to load estimates, is not accounted for in the concentration threshold.

To our knowledge, turbidity has not been employed as a surrogate for determining compliance with instream TP standards. The uncertainty in a turbidity-TP relationship may be too great for regulatory purposes, for example, at locations where dissolved phosphorus dominates. The relationships may still be valuable for understanding important periods for TP transport, and regulators can use these types of analyses in conjunction with site-specific data to determine necessary sampling frequencies. Water quality standards are typically specified with a concentration threshold and an allowable frequency of exceeding that threshold (e.g., if a criterion is only exceeded 10% of the time, the water body is not impaired KDHE, 2011). The frequency of exceeding the criterion is sufficiently high at Mendon (about 70%) to indicate that the standard is exceeded, so sampling frequency may be less important in determining compliance. At a site where the frequency of exceedence is near the acceptable level, the sampling frequency may be more important in determining compliance with the water quality standard. For example, if the acceptable frequency was 25%, it would be important to collect samples at a high frequency at Paradise to accurately determine the frequency of exceedence. The necessary degree of precision in the frequency of exceedence can also help determine minimum sampling frequency. If the frequency of exceedance needs to be determined within 5% of the true value, then weekly sampling would be adequate at both Paradise and Mendon. This type of analysis can also help regulators estimate the level of uncertainty in frequency of exceedence determined from water quality samples. At Mendon, for example, the median of the frequency of exceedence estimates was approximately 70% for monthly subsampling (Figure 10), but the spread in the distribution illustrates that it can vary by ±15%. In the absence of continuous data (e.g., at a site with similar characteristics or at the same site for a different time period), the assessment of compliance based on grab samples would likely have similar uncertainty.

Limitations of the Surrogate Approach

Previous studies have made recommendations for sampling frequency based on factors specific to study location (e.g., hydrologic characteristics), constituent of interest, and objectives of the sampling program (e.g., estimating annual loads *vs.* understanding extremes in concentration or loading). We found that the sampling frequency required to characterize annual loads of TP and TSS to within a given percentage of the reference load was site specific, even between the two sites we examined in the Little Bear River, which were very different hydrologically and had differing constituent behavior. At Paradise, the dominant form of TP was particulate, whereas at Mendon, the majority of phosphorus occurred in the dissolved form (Jones *et al.*, 2011). These differences in constituent form and behavior influence the adequacy of the surrogate relationship and drive the site-specific nature of the relationships. For example, Figure 4 shows a much higher degree of scatter in the TP-turbidity relationships at Mendon than at Paradise. Given these considerations, care should be exercised in assessing the strength of surrogate relationships. Although site specificity is a limitation of the overall surrogate approach, it is expected that the information inferred about required sampling frequency from a given site would translate to sites that have similar hydrologic and constituent behavior.

The strength of relationships between *in situ* proxies and constituents of interest determines their utility. The objectives of a monitoring program should dictate the acceptable level of uncertainty in a surrogate relationship. Uncertainty can be quantified via error propagation techniques, which account for measurement error in the sensor data, measurement error in the periodic sampled concentrations, and error in the derived surrogate relationship (Horsburgh *et al.*, 2010). Uncertainty in surrogate relationships can be expressed using prediction intervals that define a range of values for the regression estimate with a known level of certainty. Where constituent loads are to be determined, error in stage measurements, periodic discharge measurements, and resulting stage-discharge relationship must also be considered.

Although the combined uncertainty may be judged too great for some purposes (e.g., assessing compliance with numeric water quality standards), surrogate relationships can still be useful for understanding potential variability in constituent concentrations as was done in this study. All of the annual load estimates in this article were generated from concentrations derived from the same surrogate equations, and thus inherited the same level of uncertainty. The artificial subsampling isolated only the effects of reduced sampling frequency and fixed sampling timing on annual load estimates, which was our objective. Thus, the artificial subsamples represent reduced resolution observations of the natural variability in turbidity and stage, and this reduced resolution is reflected in the resulting load estimates. Given the synthetic nature of this study, we did not focus on the uncertainty in the concentration or load estimates.

SUMMARY AND CONCLUSIONS

This study used high-frequency estimates of TSS and TP concentrations derived using surrogate relationships with turbidity, along with time-matched series of discharge to calculate reference loads for two sites in the Little Bear River. To simulate decreasing sampling frequencies, the continuous records were subsampled at daily, weekly, and monthly intervals; and annual loads were calculated from the resulting subsets. At each sampling frequency for each variable and water year at each site, 10,000 realizations of annual load were generated to examine the potential variability in annual load estimates. Subsets were also created to simulate sampling at the same time every day and the same day every week. Differences between the two sites were considered in the context of their hydrologic responsiveness and dynamics of material transport. Finally, the effects of sampling frequency on the determination of compliance with a water quality concentration threshold were also examined.

Results showed that at both sites, decreasing sampling frequency increased the variability in annual load estimates; however, there was greater variability in estimated loads at the upper watershed site, and reduced sampling frequency resulted in underestimates of the reference loads at the upper watershed site, but not at the lower watershed site. These results are attributed to differences in TP and TSS sources and behavior and the flashier nature of flow and constituent transport at the upper watershed site. The time of day and the day of week that sampling was conducted were observed to have a substantial impact on estimated annual loads, although, in this case, the level of impact varied between site and year.

As the degree of variability was site specific, and depending on the acceptable error, sampling at lower frequencies may be adequate for some purposes at sites that are less flashy. Even within a watershed, different sites may require different sampling frequencies, precluding synchronized sampling programs at watershed or regional scales. However, once optimal sampling frequency is determined for a site, this information may be generalizable across sites with similar hydrologic behavior. Using a concentration-based water quality criterion, the continuous concentration estimates provided an evaluation of the uncertainty in the estimated frequency of exceedence of numeric criteria values. High-frequency measurements can be used as a tool to determine the optimal frequency for sample collection and the target locations and periods for monitoring.

ACKNOWLEDGMENTS

This study was supported by the National Science Foundation grant CBET 0610075 and by the United States Department of Agriculture's Conservation Effects Assessment Program grant UTAW-2004-05671. We gratefully acknowledge field technicians, namely Sandra Guerrero, private landowners within the Little Bear River watershed who have provided us access to their property, and Mountain Crest High School and Willow Valley Middle School for providing Internet connectivity for our remote-base stations. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the United States Department of Agriculture. Mention of trade names or commercial products within this article is for identification purposes only and does not constitute endorsement or recommendation for use by the National Science Foundation, the United States Department of Agriculture, or Utah State University.

LITERATURE CITED

Birgand, F., C. Faucheux, G. Gruau, B. Augeard, F. Moatar, and P. Bordenave, 2010a. Uncertainties in Assessing Annual Nitrate Loads and Concentration Indicators: Part 1. Impact of Sampling Frequency and Load Estimation Algorithms. Transactions of the ASABE 53:437-446.

Birgand, F., C. Faucheux, G. Gruau, F. Moatar, and M. Meybeck, 2010b. Uncertainties in Assessing Annual Nitrate Loads and Concentration Indicators: Part 2. Deriving Sampling Frequency Charts in Brittany, France. Transactions of the ASABE 54:93-104.

Brauer, N., A.T. O'Geen, and R.A. Dahlgren, 2009. Temporal Variability in Water Quality of Agricultural Tailwaters: Implications for Water Quality Monitoring. Agricultural Water Management 96:1001-1009, doi: 10.1016/j.agwat.2009.01.011.

Cassidy, R. and P. Jordan, 2011. Limitations of Instantaneous Water Quality Sampling in Surface-Water Catchments: Comparison with Near-Continuous Phosphorus Time-Series Data. Journal of Hydrology 405:182-193, doi: 10.1016/j.jhydrol.2011.05.020.

Christensen, V.G., P.P. Rasmussen, and A.C. Ziegler, 2002. Real-time Water Quality Monitoring and Regression Analysis to Estimate Nutrient and Bacteria Concentrations in Kansas Streams. Water Science and Technology 45:205-211.

Coynel, A., J. Schäfer, J.-E. Hurtrez, J. Dumas, H. Etcheber, and G. Blanc, 2004. Sampling Frequency and Accuracy of SPM Flux Estimates in Two Contrasted Drainage Basins. The Science of the Total Environment 330:233-247, doi: 10.1016/j.scitotenv.2004.04.003.

Duvert, C., N. Gratiot, J. Némery, A. Burgos, and O. Navratil, 2011. Sub-daily Variability of Suspended Sediment Fluxes in Small Mountainous Catchments – Implications for Community-Based River Monitoring. Hydrology and Earth System Sciences 15:703-713, 5, doi: 10.5194/hess-15-703-2011.

Ferguson, R.I., 1987. Accuracy and Precision of Methods for Estimating River Loads. Earth Surface Processes and Landforms 12:95-104, doi: 10.1002/esp.3290120111.

Granger, S.J., J.M.B. Hawkins, R. Bol, S.M. White, P. Naden, G. Old, G.S. Bilotta, R.E. Brazier, C.J.A. Macleod, and P.M. Haygarth, 2009. High Temporal Resolution Monitoring of Multiple Pollutant Responses in Drainage from an Intensively Managed Grassland Catchment Caused by a Summer Storm. Water, Air, and Soil Pollution 205:377-393, doi: 10.1007/s11270-009-0083-z.

Grayson, R.B., C.J. Gippel, B.L. Finlayson, and B.T. Hart, 1997. Catchment-Wide Impacts on Water Quality: The Use of "Snapshot" Sampling during Stable Flow. Journal of Hydrology 199:121-134, doi: 10.1016/S0022-1694(96)03275-1.

Horsburgh, J.S., A. Spackman Jones, D.K. Stevens, D.G. Tarboton, and N.O. Mesner, 2010. A Sensor Network for High Frequency Estimation of Water Quality Constituent Fluxes Using Surrogates. Environmental Modelling & Software 25:1031-1044, doi: 10.1016/j.envsoft.2009.10.012.

Johnes, P.J., 2007. Uncertainties in Annual Riverine Phosphorus Load Estimation: Impact of Load Estimation Methodology, Sampling Frequency, Baseflow Index and Catchment Population Density. Journal of Hydrology 332:241-258, doi: 10.1016/j.jhydrol.2006.07.006.

Jones, A.S., D.K. Stevens, J.S. Horsburgh, and N.O. Mesner, 2011. Surrogate Measures for Providing High Frequency Estimates of Total Suspended Solids and Total Phosphorus Concentrations. Journal of the American Water Resources Association 47:239-253, doi: 10.1111/j.1752-1688.2010.00505.x.

Jordan, P., J. Arnscheidt, H. McGrogan, and S. McCormick, 2005. High-Resolution Phosphorus Transfers at the Catchment Scale: The Hidden Importance of Non-storm Transfers. Hydrology and Earth System Sciences 9:685-691, doi: 10.5194/hess-9-685-2005.

Jordan, P., A. Arnscheidt, H. McGrogan, and S. McCormick, 2007. Characterising Phosphorus Transfers in Rural Catchments Using a Continuous Bank-Side Analyser. Hydrology and Earth System Sciences 11:372-381, doi: 10.5194/hess-11-372-2007.

KDHE (Kansas Department of Health and Environment, Bureau of Water), 2011. Duration and Frequency for Assessing Numeric Criteria. Kansas Department of Health and Environment, Bureau of Water, Topeka, Kansas.

Kirchner, J.W., X. Feng, C. Neal, and A.J. Robson, 2004. The Fine Structure of Water-quality Dynamics: The (High-Frequency) Wave of the Future. Hydrological Processes 18:1353-1359, doi: 10.1002/hyp.5537.

Kronvang, B. and A.J. Bruhn, 1996. Choice of Sampling Strategy and Estimation Method for Calculating Nitrogen and Phosphorus Transport in Small Lowland Streams. Hydrological Processes 10:1483-1501, doi: 10.1002/(SICI)1099-1085(199705)11:6 < 627::AID-HYP481> 3.0.CO;2-E.

Loperfido, J.V., C.L. Just, A.N. Papanicolaou, and J.L. Schnoor, 2010. *In Situ* Sensing to Understand Diel Turbidity Cycles, Suspended Solids, and Nutrient Transport in Clear Creek, Iowa. Water Resources Research 46:2-9, doi: 10.1029/2009WR008293.

Mano, V., J. Nemery, P. Belleudy, and A. Poirel, 2009. Assessment of Suspended Sediment Transport in Four Alpine Watersheds (France): Influence of the Climatic Regime. Hydrological Processes 792:777-792, doi:10.1002/hyp.7178.

Meybeck, M., L. Laroche, H.. Dürr, and J.P. Syvitski, 2003. Global Variability of Daily Total Suspended Solids and Their Fluxes in Rivers. Global and Planetary Change 39:65-93, doi: 10.1016/S0921-8181(03)00018-3.

Miller, P.S., R.H. Mohtar, and B.A. Engel, 2007. Water Quality Monitoring Strategies and Their Effects on Mass Load Calculation. Transactions of the ASABE 50:817-830.

Moatar, F. and M. Meybeck, 2007. Riverine Fluxes of Pollutants: Towards Predictions of Uncertainties by Flux Duration Indicators. Comptes Rendus Geosciences 339:367-382, doi: 10.1016/j.crte.2007.05.001.

Neilson, B.T. and S.C. Chapra, 2003. Integration of Water Quality Monitoring and Modeling for TMDL Development. Water Resources Impact 5:9-11.

Pappas, E.A. and C. Huang, 2008. Policy Analysis Predicting Atrazine Levels in Water Utility Intake Water for MCL Compliance. Environmental Science & Technology 42:10-14, doi: 10.1021/es800457v.

Phillips, J.M., B.W. Webb, D.E. Walling, and G.J.L. Leeks, 1999. Estimating the Suspended Sediment Loads of Rivers in the LOIS Study Area Using Infrequent Samples. Hydrological Processes 13:1035-1050, doi: 10.1002/(SICI)1099-1085(199905)13:7 < 1035::AID-HYP788 > 3.0.CO;2-K.

Quilbe, R., A. Rousseau, M. Duchemin, A. Poulin, G. Gangbazo, and J. Villeneuve, 2006. Selecting a Calculation Method to Estimate Sediment and Nutrient Loads in Streams: Application to the Beaurivage River (Québec, Canada). Journal of Hydrology 326:295-310, doi: 10.1016/j.jhydrol.2005.11.008.

Richards, R.P. and J. Holloway, 1987. Monte Carlo Studies of Sampling Strategies for Estimating Tributary Loads. Water Resources Research 23:1939, doi:10.1029/WR023i010p01939.

Robertson, D.M., 2003. Influence of Different Temporal Sampling Strategies on Estimating Total Phosphorus and Suspended Sediment Concentration and Transport in Small Streams. Journal of the American Water Resources Association 39:1281-1308, doi: 10.1111/j.1752-1688.2003.tb03709.x.

Robertson, D.M. and E.D. Roerish, 1999. Influence of Various Water Quality Sampling Strategies on Load Estimates for Small Streams. Water Resources Research 35:3747-3759, doi: 10.1029/1999WR900277.

Schleppi, P., P.A. Waldner, and B. Fritschi, 2006. Accuracy and Precision of Different Sampling Strategies and Flux Integration Methods for Runoff Water: Comparisons Based on Measurements of the Electrical Conductivity. Hydrological Processes 20:395-410, doi:10.1002/hyp.6057.

Scholefield, D., T. Le Goff, J. Braven, L. Ebdon, T. Long, and M. Butler, 2005. Concerted Diurnal Patterns in Riverine Nutrient Concentrations and Physical Conditions. The Science of the Total Environment 344:201-10, doi: 10.1016/j.scitotenv.2005.02.014.

Soil Survey Staff NRCS USDA, 2008. Web Soil Survey, http://websoilsurvey.nrcs.usda.gov/.

Stelzer, R.S. and G.E. Likens, 2006. Effects of Sampling Frequency on Estimates of Dissolved Silica Export by Streams: The Role of Hydrological Variability and Concentration-Discharge Relationships. Water Resources Research 42:1-10, doi: 10.1029/2005WR004615.

Tomlinson, M.S. and E.H. De Carlo, 2003. The Need for High Resolution Time Series Data to Characterize Hawaiian Streams. Journal of the American Water Resources Association 39:113-123, doi: 10.1111/j.1752-1688.2003.tb01565.x.

Utah Administrative Code R317-2, 2011. Standards of Quality for Waters of the State. November 1, 2011.

Utah DEQ (Utah Department of Environmental Quality, TMDL Section), 2000. Little Bear River Watershed TMDL. Utah Department of Environmental Quality, TMDL Section, Salt Lake City, Utah.

de Vries, A. and H.C. Klavers, 1994. Riverine Flues of Pollutants: Monitoring Strategy First, Calculation Methods Second. European Water Pollution Control 4:12-17.

Walling, D.E. and W. Webb, 1981. The Reliability of Suspended Sediment Load Data. Proceedings of the Florence Symposium, IAHS, International Association of Hydrological Sciences, Washington, D.C., pp. 177-194.

Webb, B., J. Phillips, and D. Walling, 2000. A New Approach to Deriving "Best-Estimate" Chemical Fluxes for Rivers Draining the LOIS Study Area. The Science of the Total Environment 251-252:45-54, doi: 10.1016/S0048-9697(00)00413-7.

Whyte, D.C. and J.W. Kirchner, 2000. Assessing Water Quality Impacts and Cleanup Effectiveness in Streams Dominated by Episodic Mercury Discharges. The Science of the Total Environment 260:1-9, doi: 10.1016/S0048-9697(00)00537-4.