Comparison of methods for estimating sediment and nitrogen loads from a small agricultural watershed

A. Zamyadi[1], J. Gallichand[1]* and M. Duchemin[2]

[1]*Département des Sols et de Génie Agroalimentaire, Université Laval, Québec, Québec G1V 0A6, Canada; and* [2]*Institut de Recherche et de Développement en Agroenvironnement Inc., Québec, Québec G1P 3W8, Canada.* **Email: jacques.gallichand@sga.ulaval.ca*

Zamyadi, A., Gallichand, J. and Duchemin, M. 2007. **Comparison of methods for estimating sediment and nitrogen loads from a small agricultural watershed**. Canadian Biosystems Engineering/Le génie des biosystèmes au Canada **49**: 1.27-1.36. The knowledge of nutrient mass load at the outlet of watersheds is a key tool in water quality management projects. However, because of the lack of frequent concentration measurements, a precise estimation of mass load is not possible. This study was conducted to determine the quality of mass load estimation for a combination of six sampling frequencies (daily to monthly), two sampling schemes (fixed and random), and seven calculation methods (averaging, ratio estimator, and interpolation) for sediment and three nitrogen components at the outlet of a 5.3 km^2 agricultural watershed in the province of Québec. Hourly values of flow, total nitrogen, nitrate, ammonium, and sediment were generated for a two-year period by the HSPF model calibrated for the watershed. Load estimates based on the mass accumulation of hourly values were assumed to represent the true loads to which loads estimated by other methods were compared using bias, standard deviation, and root mean square error (RMSE). For all water quality parameters, the RMSE decreased with the sampling frequency. Fixed sampling schemes always resulted in RMSE values less than those of random schemes. This better performance of the fixed sampling scheme is attributed to the autocorrelation in time for all water quality parameters. The sediment autocorrellogram showed a 24-hour periodicity that is explained by snowmelt and more frequent evening and night rainfalls. All load evaluations generally resulted in an underestimation of the true load, most likely due to the flashy hydrological response of the small watershed studied. Of the seven load calculation methods studied, linear interpolation, which is used to obtain an estimate of concentration for each available hourly flow value, systematically yielded the lowest RMSE value. However, ratio estimator methods did not fare well, ranking only 5 and 6 out of the seven methods tested. **Keywords:** load estimation, water sampling, nitrogen, sediment, agricultural watershed.

La charge en nutriments à l'exutoire d'un bassin versant est un élément-clé pour la gestion de la qualité de l'eau. Toutefois, à cause du peu de mesures de concentration, un estimé précis de la charge n'est pas possible. Cette étude a été effectuée pour déterminer la qualité d'estimation de la charge pour une combinaison de six fréquences d'échantillonnage (journalière à mensuelle), deux schémas (fixe et aléatoire) et sept méthodes de calcul (moyenne, rapport d'estimation, interpolation) pour les sédiments et trois composantes de l'azote pour un bassin versant de 5.3 km^2 du Québec. Des valeurs horaires de débit, azote total, nitrate, ammonium et sédiment ont été générées pour une période de deux ans avec le programme HSPF calibré pour ce bassin versant. Il a été supposé que les estimés de charge basés sur l'accumulation massique des valeurs horaires représentent les valeurs réelles auxquelles les valeurs estimées par les autres méthodes peuvent être comparées en utilisant le biais, l'écart type et la racine carrée de l'erreur moyenne au carré (RMSE).Pour tous les paramètres de qualité de l'eau, RMSE diminue avec la fréquence d'échantillonnage. L'échantillonnage fixe résulte toujours en un RMSE plus faible que pour l'échantillonnage aléatoire; ce qui peut être attribué à l'autocorrellation temporelle. L'autocorrellogramme des sédiments en suspension montre une périodicité de 24 heures, expliquée par la fonte nivale printanière et par les précipitations plus fréquentes durant la soirée et la nuit. Les évaluations de la charge ont presque toujours résulté en une sous-estimation, probablement due aux crues de courte durée caractérisant ce bassin versant. Des sept méthodes de calcul étudiées, l'interpolation linéaire a systématiquement donné des RMSE plus faibles. Toutefois, les méthodes du rapport d'estimation (incluant la méthode de Beale) n'ont pas bien performé, obtenant la cinquième et sixième place. **Mots-clés:** estimation, charge polluante, échantillonnage, eau, azote, sédiments, bassin versant.

INTRODUCTION

Agricultural pollutants can degrade surface water and result in eutrophication and deterioration of drinking water quality. Watershed-scale management programs have been proven to be efficient in reducing water pollution from agricultural activities, but their implementation requires an estimation of the pollutant load at the watershed outlet (Guo et al. 2002; Gangbazo et al. 1994). The estimated load can be used in pollutant budget calculations and to detect trends in water pollution. Among agricultural pollutants, phosphorus, nitrogen, and sediment contribute to the deterioration of aquatic environment recreational potential and to the increase of water treatment costs (Harmel et al. 2003; Legeas and Iachkine 1992). Pollution of surface water by sediment and nitrogen is a serious problem in southern Québec, particularly in the agricultural watersheds of the Châteauguay, Yamaska, Boyer, and Chaudière rivers (Hébert and Ouellet 2005; Painchaud 1996).

The pollutant true load at the outlet of a watershed for a given period of time can never be known exactly, but can be estimated with reasonable precision by integrating the product of measured concentrations and corresponding flows for short time intervals (Littlewood 1992). The use of automated equipment allows precise and economical flow measurement for short time intervals (e.g. hourly or less), but the measurement of a pollutant concentration requires water sampling, storing, and costly laboratory analyses, which makes concentration measurement the limiting factor for estimating pollutant loads

(Moatar and Meybeck 2005; Stone et al. 2000). Assuming that flow measurement is not limiting, the accuracy of load estimation depends on the sampling method (i.e. discrete or composite), the sampling frequency, the sampling scheme (i.e. random or fixed), and the load calculation method (Cohn 1995).

Composite sampling involves the collection of several flow- or time-proportional aliquots in a single sampling bottle. Composite sampling may give an accurate estimation of a contaminant load, but requires automated and usually costly sampling equipment (Stone et al. 2000). Unlike composite sampling, discrete sampling, often referred to as grab sampling, does not require sophisticated and expensive equipment and trained personnel. Discrete sampling usually consists of the manual collection of water samples at a predetermined frequency. Because of its simplicity, discrete sampling has been used in many water quality monitoring programs in North America and Europe (Swistock et al. 1997). In the United States, discrete sampling has been adopted by several state agencies (Martin et al. 1992; Robinson et al. 2004).

A sampling program is usually specified by its sampling frequency and its sampling scheme. The sampling frequency corresponds to the time interval between two consecutive samples and the sampling scheme may be fixed (e.g. every Monday of a month for a monthly frequency) or random (e.g. any day of the month for a monthly frequency). In a random scheme, the time interval between two consecutive samples is, on average, equal to the sampling frequency. Widely used sampling frequencies are monthly, bimonthly (approximately one time per two weeks) and weekly (Williams et al. 2004). These frequencies are also recommended by the Québec Ministry of Environment (Hébert and Légaré 2000). Since flow and contaminant concentration vary less for large watersheds, the optimal sampling frequency is less for large watersheds than small watersheds (Tate et al. 1999; White 1999). Attempts to compare fixed and random sampling strategies have been limited mostly to large watersheds (King and Harmel 2003; Wang et al. 2003).

According to Moatar and Meybeck (2005) and Preston et al. (1989), load calculation methods can be classified as averaging, ratio estimator, interpolation, and regression. Averaging, the product of mean concentration by mean flow is simple, flexible, and easy to calculate, but an estimation bias is inevitable if data are not collected from the entire range of flows and concentrations (Ferguson 1987; Dolan et al. 1981). The ratio estimator weights concentration with the corresponding flow at sampling times. Beale (1962) developed an improved ratio estimator that incorporates the covariance structure between load and flow in order to reduce the estimation bias. Although developed for large watersheds, ratio estimator methods have also proven accurate for estimation of nitrogen and phosphorus loads from watersheds less than 20 km^2 in England (Littlewood 1995) and Finland (Rekolainen et al. 1991).

In their simplest forms, interpolation and regression methods aim at filling the pollutant concentration gaps between measured flow values to obtain high resolution flow – concentration time series that will be used to estimate the true load (Sivakumar and Wallender 2004). In the case of interpolation, only concentration data are used, whereas in regression, independent variables can be flow, time of the year, electrical conductivity, pH, etc. (Preston et al. 1989). The most frequently used interpolation is linear (Moatar and Meybeck 2005). It has been used to calculate loads of nitrogen and phosphorus from tributaries of the Baltic Sea (Stålnacke et al. 1999) and of sediment in various rivers of France (Toma et al. 1993). Instead of predicting concentration, regression methods are often used to predict the load at unsampled concentration points (Haggard et al. 2003). The regression method known as the rating curve is essentially a loglinear regression model between load and flow (Cohn et al. 1992). The concentration of highly mobile chemicals, like nitrate, is not well correlated with flow, so the use of regression is not recommended for load estimation in this case (Robertson and Roerish 1999).

According to Phillips et al. (1999), the efficiency of a water sampling strategy and calculation method, in estimating the true load of a pollutant, is usually assessed in terms of accuracy (low bias) and precision (low dispersion). For nutrient and sediment loads, King and Harmel et al. (2003) found that a 60-minute flow and concentration sampling interval would result in an estimated load within 7% of the true load. This high frequency sampling is not realistic in practical monitoring situations. To study the efficiency of load estimation strategies, procedures have been developed to generate virtual high frequency concentration and flow time series. In the Great Smoky Mountains National Park in the eastern USA, Robinson et al. (2004) used a multiple linear regression model to generate high frequency concentrations of nitrate that were used to test the effect of various sampling frequencies. To compare the accuracy of discrete and composite sampling, Whelan et al. (1999) used a water quality model to generate concentrations of different ions in the Lambro River in northern Italy. Synthetic concentration time series have been generated by Littlewood (1995) using a first-order transfer function between concentration and flow to test sampling frequencies and calculation methods. Hydrologic water quality models have also been used for generating high frequency series to test sampling strategies. Vandenberghe et al. (2002) used the Soil Water and Assessment Tool (SWAT) to generate hourly flow and water quality values to evaluate sampling strategies. Detailed hydrological-water quality models, such as the Hydrological Simulation Program-Fortran (HSPF) model, can provide accurate estimates of sediment and nutrient concentrations (USEPA 2003). HSPF is considered one of the most exhaustive models for water quality simulation (Bergman et al. 2002; Bicknell et al. 1993; Donigian and Hubert 1991). HSPF has been successfully validated and verified for flow and water quality for agricultural watersheds in Iowa (Donigian et al. 1995), Tennessee (Chew et al. 1991), France (Kauark Leite 1990), and in Québec by Laroche et al. (1996) and Bernier and Gallichand (1999). Bergman et al. (2002) have observed that HSPF is increasingly used for water quality assessment studies involving sediment, phosphorus, and nitrogen.

Systematic examination of different load calculation methods in combination with different sampling strategies has been rarely done for nitrogen constituents and sediment on small watersheds. The objective of this research was to determine the best load estimation strategy for nitrogen constituents and sediment at the outlet of a 5.3 km^2 watershed in southern Québec. Load estimation strategies were formed by a combination of six sampling frequencies, two sampling schemes, and seven estimators, and used two years of hourly time series generated by HSPF.

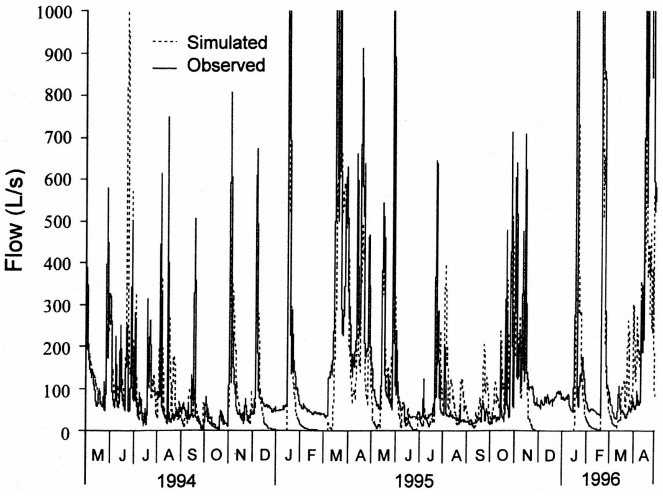

Fig. 1. Observed and HSPF simulated mean daily flow.

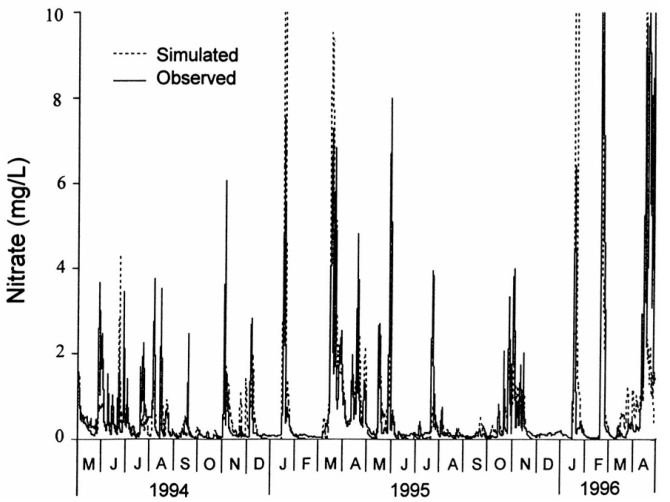

Fig. 2. Observed and HSPF simulated mean daily nitrate concentration.

MATERIALS and METHODS

Time series generation

Version 10.1 of HSPF was calibrated and verified by Bernier and Gallichand (1999) for a two-year period (May 1994 to April 1996) for flow, total nitrogen, nitrate, ammonium, and sediment at the outlet of the Turmel Brook watershed, by running the model continuously on a hourly time step. The watershed was divided into 48 pervious land segments and each segment was assigned its specific physical characteristics and land use corresponding to the two years of monitoring. Input time series consisted of hourly air and dew point temperatures, total precipitation, potential evapotranspiration, wind speed, and solar radiation. In the present study, we used the calibrated model output time series from which hourly masses of sediment and nitrogen components were accumulated and assumed to represent the true load of each component. A two-year period is considered sufficient for the study of water sampling strategies (Guo et al. 2002; Rekolainen et al. 1991). Figures 1 and 2 present observed and HSPF simulated flow and nitrate, respectively, where the ordinates have been truncated to improve legibility. Because the actual sampling frequency ranged from 3 to 12 samples per week, the observed values of Figs. 1 and 2 were generated with the interpolation function of HSPF. Observed and simulated flows at the outlet of the watershed (Fig. 1) follow the same pattern with the exception that simulated values tend to be less than observed ones during low flow periods. This small deviation is not deemed important, in view of sampling strategies evaluation, because most pollutants are transported during periods of high flow. A good agreement is also observed in the case of nitrate (Fig. 2) and for the other water quality parameters. Examination of Figs. 1 and 2 shows that HSPF simulated time series represent a realistic situation which can be used in the study of water sampling strategies for load determination.

Watershed description and data measurement

The Turmel Brook watershed is located near Ste-Marie de Beauce, Québec, about 60 km south of Québec City. Its average slope is 2.4%, elevation 303 m above mean sea level, and total area 5.3 km^2. At the time of measurements (1994-1996), the watershed area consisted of 11.9% of pasture, 46.3% of hay fields, 1.2% of cereals, 3.2% of fallow, and 37.4% of woodland. The Turmel Brook is a tributary of the Bélair River which is a supplemental source of drinking water for the city of Sainte-Marie. Soils are primarily silty loams, rich in organic material, over sandy to clay loams. Soil permeability is classified as low to very low and the depth of soil over bedrock ranges from 2 to 5 m. Normal climatic values (1961-1990) were 4.3°C for temperature, 1058 mm for total precipitation, with about 20% in snow, and 500 mm for potential evapotranspiration (Environment Canada 1993). Fertilization was almost exclusively from manure applications (Chokmani and Gallichand 1997).

A gauging station, at the outlet of the study watershed, consisted of an insulated shelter, a semi-circular control section, and equipment for measurement of flow and water sampling. The station was in operation from March 1994 to October 1996, with records available from May 1994 to April 1996 on a continuous basis. For these two years, observed average values were 127.9 L/s for flow, 0.865 mg/L for total nitrogen, 0.204 mg/L for ammonium, 0.660 mg/L for nitrate, and 0.012 mg/L for sediment. The presence of nitrogen in the surface waters of the watershed was due to manure applications in excess of crop requirements. Sediment and nitrogen components (total nitrogen, nitrate, and ammonium) have been examined in the surface waters because of possible problems related to drinking water supplies. In many areas of the study watershed, slopes are steep and erosion was visible (Gallichand et al. 1998). Sediment transported by surface water accumulates in water bodies, can cause operational problems for water treatment and distribution equipment, and their removal can be very costly (Peart 1995; Sundborg and Rapp 1986). Nitrate in drinking water can cause methemoglobinemia in young children, stomach cancer in adults, and poison livestock (Chambers et al. 2002; Tebbutt 1998).

Sampling strategy simulations

The sampling strategies tested are based on the widely used discrete sampling method. All possible combinations of six sampling frequencies (monthly, bimonthly, weekly, biweekly (two times per week), triweekly (three times per-week), and

daily), two sampling schemes (fixed and random intervals), and seven calculation methods (numbered from M1 to M7) resulted in 84 sampling strategies. All flow and concentration data needed for the evaluation of each sampling strategy consisted of a subset of the complete two-year hourly time series generated by HSPF. For the fixed interval scheme, a sample consisting of a water quality parameter concentration and the corresponding flow was drawn from the complete time series at a time interval corresponding to the desired frequency. For a given frequency, all possible realizations were sampled and included in the analyses. For example, a weekly sampling frequency resulted in 45 realizations (i.e. 8:00 to 16:00 from Monday to Friday). For the random scheme, sampling for each realization was done on a replacement basis assuming a uniform distribution over the same sampling time extent used for the fixed scheme (e.g. from 8:00 to 16:00 for a daily interval). For comparison purposes, the number of realizations used for the random schemes was the same as that for the fixed schemes, and varied from nine for a daily sampling frequency to 135 for a monthly frequency.

Calculation methods

Of the numerous methods available to estimate the average load of a water quality parameter, seven were tested. The choice of methods has been restricted to those that can be computed without involved statistical analyses. For this reason, regression methods have been left out and only averaging, ratio estimator, and linear interpolation methods were kept. Of the seven methods selected, and described below, only methods M1 and M2 do not use all flow measurements.

Method 1 (M1) is the product of the mean sampled concentration by the mean sampled flow for the period of interest.

$$\hat{L}_1 = \left(\sum_{n=1}^{N} \frac{C_n}{N}\right)\left(\sum_{n=1}^{N} \frac{Q_n}{N}\right) \tag{1}$$

where:
$\hat{L}_1$ = average estimated load (M/T) for method M1,
C = instantaneous concentration (M/L^3),
Q = instantaneous flow (L^3/T),
N = number of concentration and flow measurements, and
n = an index for N.

Method 2 (M2) evaluates the load as the mean of all sampled instantaneous loads during a given period.

$$\hat{L}_2 = \sum_{n=1}^{N} \frac{C_n \times Q_n}{N} \tag{2}$$

where: $\hat{L}_2$ = average estimated load (M/T) for method M2.

Method 3 (M3) is a backward-flow load estimator. It uses, as flow value, the mean of all available flows bounded by two consecutive concentrations corresponding to sampling times t_n and t_{n-1}.

$$\hat{L}_3 = \sum_{n=1}^{N} \left(C_n \overline{Q_{n,n-1}}\right) \tag{3}$$

where:
$\hat{L}_3$ = average estimated load (M/T) for method M3 and
$\overline{Q}_{n,n-1}$ = mean of measured flow values (L^3/T) between times t_n and t_{n-1}.

The flow measurements used to determine the mean flow value are normally much more numerous than the concentration measurements.

Method 4 (M4) is the product of the mean concentration by the mean flow for the complete period of interest.

$$\hat{L}_4 = \overline{Q}\left(\sum_{n=1}^{N} \frac{C_n}{N}\right) \tag{4}$$

where:
$\hat{L}_4$ = average estimated load (M/T) for method M4 and
$\overline{Q}$ = average of all flow measurements (L^3/T) that may or may not coincide with a concentration measurement.

Averaging methods M1 to M4 is often used for load estimation (Moatar and Meybeck 2005; Littlewood 1995).

Method 5 (M5) is the ratio estimator, i.e. the product of the flow-weighted mean concentration by the mean flow for the complete period of measurement.

$$\hat{L}_5 = \frac{\sum_{n=1}^{N} C_n Q_n}{\sum_{n=1}^{N} Q_n} \times \overline{Q} \tag{5}$$

where: $\hat{L}_5$ = average estimated load (M/T) for method M5.

Method 6 (M6) is a modification of the ratio estimator suggested by Beale (1962) to correct the bias introduced by Eq. 5. This method is suited to situations where the number of flow measurements is large compared to the number of concentration measurements.

$$\hat{L}_6 = \hat{L}_5 \left[\frac{1 + \frac{Cov(Q_n, L_n)}{N\overline{Q}_* \overline{L}_*}}{1 + \frac{Var(Q_n)}{N(\overline{Q}_*)^2}}\right] \tag{6}$$

where:
$\hat{L}_6$ = average estimated load (M/T) for method M6,
L_n, Q_n = measured load and flow values,
$\overline{L}_*$, $\overline{Q}_*$ = means of all measured loads and flows,
$Cov(Q_n,L_n)$ = covariance between flow and load at the moments of sampling, and
$Var(Q_n)$ = variance of measured flow values at the moment of sampling.

Method 7 (M7) uses a linear interpolation between two consecutive measured concentrations to obtain a value for each measured flow value.

$$\hat{L}_7 = \sum_{m=1}^{M} \frac{C_m^{\text{int}} Q_m}{M} \tag{7}$$

where:
$\hat{L}_7$ = average estimated load (M/T) for method M7,
C_m^{int} = interpolated concentration corresponding to the m^{th} flow value,
Q_m = a measured flow value, and
M = total number of all measured flow values.

Table 1. Effect of the sampling frequency on load estimation statistics.

Water quality parameters	Sampling frequency	Bias (kg/h)	Standard deviation (kg/h)	RMSE (kg/h)
Total nitrogen	daily	-0.85808	1.10702	1.40066 a*
	triweekly	-0.92265	1.20942	1.52119 a
	biweekly	-0.94236	1.44924	1.72868 b
	weekly	-0.93284	2.12108	2.31716 c
	bimonthly	-1.02111	2.95405	3.12554 d
	monthly	-0.74928	4.48007	4.54229 e
Nitrates	daily	-0.68758	0.89716	1.13035 a
	triweekly	-0.74274	0.98793	1.23597 a
	biweekly	-0.75779	1.19088	1.41156 b
	weekly	-0.74565	1.75439	1.90629 c
	bimonthly	-0.81588	2.44976	2.58205 d
	monthly	-0.59236	3.70925	3.75625 e
Ammonium	daily	-0.17178	0.20201	0.26518 a
	triweekly	-0.18080	0.21185	0.27852 a
	biweekly	-0.18554	0.24546	0.30769 b
	weekly	-0.18926	0.34462	0.39316 c
	bimonthly	-0.20725	0.47236	0.51582 d
	monthly	-0.16137	0.72317	0.74096 e
Sediment	daily	-0.00217	0.01787	0.01800 a
	triweekly	-0.00218	0.02082	0.02094 b
	biweekly	-0.00269	0.02501	0.02516 c
	weekly	-0.00259	0.03619	0.03628 d
	bimonthly	-0.00457	0.05048	0.05069 e
	monthly	0.00011	0.08148	0.08148 f

*RMSE values with the same letter are not significantly different at the 95% probability level.

Data analysis

As suggested by, among others, Haggard et al. (2003) and Guo et al. (2002), the calculated loads were compared to the true load using the root mean square error (RMSE):

$$RMSE = \sqrt{\bar{\varepsilon}^2 + s^2} = \sqrt{\left(\bar{L} - L_o\right)^2 + \frac{\sum_{i=1}^{N}\left(L_i^* - \bar{L}\right)^2}{N}} \qquad (8)$$

where:

$\bar{\varepsilon}$ = bias,
s = standard deviation,
$\bar{L}$ = average of estimated loads,
L_o = true load, and
L_i^* = individual estimated load values.

The statistic RMSE incorporates an estimation of accuracy (i.e. bias) and precision (i.e. standard deviation). The value of RMSE compares the average of estimated loads to the true load, and individual estimated load values to their average. Since the experimental setup includes all possible realizations for any combination of the three experimental parameters (i.e. six sampling frequencies, two sampling schemes, and seven calculation methods), a total of 6426 values were calculated for each of the four water quality parameters (sediment, total nitrogen, ammonium, and nitrate). The value of N in Eq. 8 ranged from 126 to 3213 depending on the combination to which it applied. Each component of an experimental parameter (e.g. each of the six sampling frequencies) was ordered from lowest to highest RMSE, and each RMSE was assigned a 95% confidence interval using the bootstrap technique with 5000 replacement subsamples (Efron and Tibshirani 1993). Therefore, two components of an experimental parameter do not differ from one another, at the 95% level, if their confidence intervals overlap.

RESULTS and DISCUSSION

Sampling frequency

For each water quality parameter, Table 1 presents the sampling frequencies ordered according to increasing RMSE values. For all water quality parameters, the RMSE value increases with the time interval between two successive water samplings. Values of RMSE in Table 1 clearly show better load estimation for higher sampling frequencies resulting from better information about the temporal variation of flow and concentration. This observation is supported by many studies on load estimation of water quality parameters (Miller et al. 2000) and, in particular, for nitrate load at the outlet of large agricultural watersheds (Guo et al. 2002).

Additional understanding on the effect of sampling frequency on load estimation can be gained by examining the two components of RMSE, i.e. bias and standard deviation. For all sampling frequencies and water quality parameters, the standard deviation of estimation follows very closely that of the RMSE, which indicates that standard deviation is an important component of RMSE. Standard deviation decreases for higher sampling frequency. However, bias follows the same trend as that of RMSE only from daily to bimonthly frequencies. From bimonthly to monthly frequencies, bias decreases sharply for all four water quality parameters of Table 1, to a point where the bias is always less for the monthly frequency than for the daily frequency. A similar behavior has been found for loads of nitrate and total solutes by Fogle et al. (2003) for a small watershed in central Kentucky. Although the bias value for the monthly frequency is the smallest, the corresponding standard deviation is the largest. A small bias will not result in a better load estimate if the standard deviation is large, and vice versa. Dolan et al. (1981) noted that it is therefore required to make a tradeoff between bias and standard deviation by using a single criterion to evaluate the efficiency of load estimation strategies.

The lower bias value for the monthly frequency might be explained, at least partly, by the number of realizations which varies from one frequency to another. The large time interval between monthly samples results in 1890 realizations to cover all possible fixed and random sampling possibilities. For the daily frequency, the total number of realizations is only 126. However, daily and monthly sampling frequencies biases are very close (e.g. -0.17178 vs -0.16137 kg/h, respectively, for ammonium) because the daily frequency compensates the smaller number of realizations by a larger number of samples taken during the two-year study period. This pattern of the bias

Table 2. Effect of sampling scheme on load estimation statistics.

Water quality parameters	Sampling scheme	Bias (kg/h)	Standard deviation (kg/h)	RMSE (kg/h)
Total nitrogen	fixed	-1.12187	2.15074	2.42576 a*
	random	-0.66725	3.58430	3.64588 b
Nitrates	fixed	-0.90808	1.77102	1.99025 a
	random	-0.52194	2.97047	3.01598 b
Ammonium	fixed	-0.21249	0.36401	0.42149 a
	random	-0.15121	0.57218	0.59183 b
Sediment	fixed	-0.00627	0.03986	0.04036 a
	random	-0.00204	0.06301	0.06305 b

*RMSE values with the same letter are not significantly different at the 95% probability level.

increasing, from daily to bimonthly samplings, and then of decreasing from bimonthly to monthly samplings, cannot be attributed solely to the number of realizations since it has been reported by Guo et al. (2002) for nitrate and by Robertson and Roerish (1999) for total phosphorus, for completely different experimental settings.

Results of the confidence interval analyses, shown as letters along the RMSE values in Table 1, indicate that sediment behaves differently than the three nitrogen constituents. For sediment, all sampling frequencies result in significantly different RMSE values, whereas for total nitrogen, nitrate, and ammonium, the daily and triweekly frequencies are not significantly different from each other. Therefore, for this watershed and the three nitrogen constituents considered, there are no advantages in sampling water daily instead of three times a week.

Sampling scheme

In Table 2, the sampling schemes are classified by increasing RMSE values for each of the water quality parameters. The confidence interval analysis on RMSE shows that fixed and random samplings are significantly different for all parameters. Fixed sampling always resulted in lower RMSE values and lower standard deviations; an observation that can be explained by the presence of autocorrelation in time for all of the water quality parameters. According to Cochran (1977), a fixed sampling plan will be more precise than a random one for a time series with autocorrelation. Autocorrelation in time was present for all water quality parameters and shown in Fig. 3 for total nitrogen and sediment using the complete hourly time series. The autocorrellograms of nitrate and ammonium were very similar to that of total nitrogen. For sediment concentration, Fig. 3 shows a periodicity of 24 hours which can be explained partly by the daily cycle of spring snowmelt and partly by the systematic higher incidence of rainfall observed between 20:00 and 2:00. These two factors result in an increase of flow rate and amount of sediment in suspension. That phenomenon is not observed for nitrogen constituents because they are mainly in a dissolved state and less affected by changes in flow magnitude.

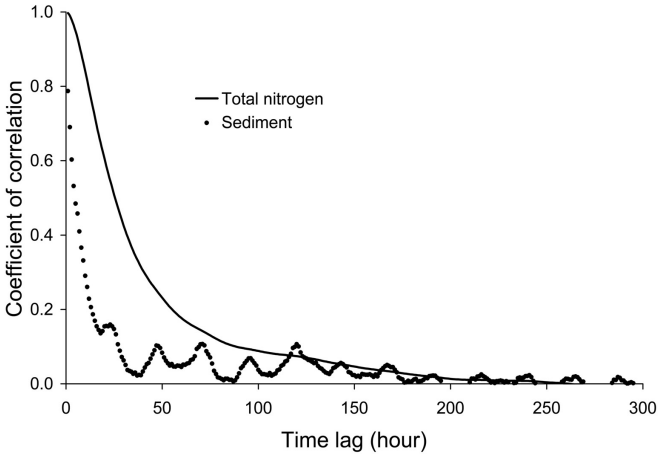

Fig. 3. Time correllogram for two water quality parameters.

Although the load estimation by the fixed scheme is always more precise, it is also less accurate, as can be seen by the more important bias in Table 2. Biases in Tables 1 and 2 are almost always negative, indicating underestimation. The preponderance of negative bias can be explained by the flashy response of the watershed resulting in few, but important, load peaks; a behavior often observed for small watersheds (Baker et al. 2004). This underestimation of the load is consistent with the percentage of the two-year period with an hourly load below the true load, which ranges from 89% for ammonium to 93% for sediments. This high probability of an hourly load below the true load explains the negative bias.

Calculation method

The seven methods are presented in order of increasing RMSE (Table 3) for each of the water quality parameters. Whatever the water constituent considered, the ranking of the methods in increasing RMSE order, is always the same: 7, 3, 1, 4, 5, 6, and 2. A similar ranking was expected for the three nitrogen constituents, but not for sediment. Linear interpolation, used to obtain an estimate of concentration for each available hourly flow value (M7), is systematically the best load approximation method. However, when taking bias and standard deviation separately, linear interpolation rarely give the lowest values. Linear interpolation for nitrate load calculation has been found the best performing method by Moatar and Meybeck (2005). Also, Kronvang and Bruhn (1996) found that, for two watersheds of 8.5 and 103 km^2 in East Denmark, linear interpolation gave the best estimation of total nitrogen annual transport and concluded it should be routinely used by Danish water quality monitoring programs.

Method 2 is the simplest to use, but contrarily to method 7, it gave the worst load estimation. That poor performance is related to ignoring flow values other than those taken at the time of water sampling, resulting in a poor coverage of flow events (Walker 1999).

Although ratio estimator methods fared well for estimating nitrogen loads from small watersheds in England (Littlewood 1995) and Finland (Rekolainen et al. 1991), the ratio estimator

Table 3. Effect of the calculation method on load estimation statistics.

Water quality parameters	Calculation method	Bias (kg/h)	Standard deviation (kg/h)	RMSE (kg/h)	
Total nitrogen	M7	-1.26381	0.91439	1.55991	a*
	M3	-1.81176	0.53348	1.88867	b
	M1	-2.07743	0.70838	2.19487	c
	M4	-2.31312	0.17059	2.31940	d
	M5	0.12467	2.96392	2.96653	e
	M6	0.34967	3.41007	3.42795	e
	M2	0.72985	5.46235	5.51090	f
Nitrates	M7	-1.01161	0.75512	1.26238	a
	M3	-1.47350	0.39661	1.52596	b
	M1	-1.67550	0.54827	1.76291	c
	M4	-1.85909	0.13454	1.86396	d
	M5	0.11091	2.47091	2.47339	e
	M6	0.29644	2.84348	2.85890	e
	M2	0.60727	4.53170	4.57221	f
Ammonium	M7	-0.25102	0.15179	0.29334	a
	M3	-0.33093	0.13950	0.35914	b
	M1	-0.39551	0.16365	0.42802	c
	M4	-0.44674	0.03619	0.44819	d
	M5	0.00392	0.45753	0.45754	d
	M6	0.04116	0.52517	0.52677	d
	M2	0.10614	0.86630	0.87278	e
Sediment	M7	-0.00659	0.01677	0.01802	a
	M3	-0.02305	0.00374	0.02335	b
	M1	-0.01846	0.01458	0.02352	b
	M4	-0.01994	0.01643	0.02584	b
	M5	0.01322	0.05432	0.05591	c
	M6	0.01674	0.06234	0.06455	c
	M2	0.02327	0.09842	0.10113	d

*RMSE values with the same letter are not significantly different at the 95% probability level.

(M5) and the Beale's ratio estimator (Beale 1962, M6) ranked only fifth and sixth best methods based on the RMSE (Table 3), and were not significantly different from one another. Comparing different load calculation methods, Rekolainen et al. (1991) found that ratio estimator methods (M5 and M6) yielded low biases, but at the cost of high standard deviations. The low biases observed in Table 3 for M5 and M6 are explained by their meeting two conditions (Preston et al. 1989): 1) the regression between flow and load gives a straight line passing through the origin (R^2 ranged from 0.3 to 0.6), and 2) the variance of the load is proportional to flow. However, these conditions do not guarantee accuracy, as can be seen in Table 3. Although Beale's method was developed to reduce the bias of the ratio estimator method, method M6 always resulted in larger biases and standard deviations than method M5. Despite the good performance on large and midsize watersheds (Cohn 1995), Beale's method does not appear to have been verified for small watersheds where streamflow fluctuations are important and for which the processes of delivery, retention, and resuspension of nutrients operate differently (Kronvang and Bruhn 1996).

Table 3 shows that methods 1, 3, 4, and 7 underestimate the true load, whereas methods 2, 5, and 6 overestimate it. Sediment behaves differently from the three nitrogen constituents with method 7 giving the lowest bias compared to methods 6 and 5 for nitrogen. Highly biased methods should be avoided because the estimate will be in error no matter how many samples are taken (Dolan et al. 1981). Therefore, methods 1 and 4 should be avoided for load estimation in the experimental watershed.

Number of sampling realizations and load estimation

To evaluate the effect of sampling schemes and calculation methods, we used a number of realizations that varied only with the sampling frequency. Each realization results in a different calculated load. Figure 4 shows, for total nitrogen, the range of load values obtained for the six sampling frequencies, the two schemes, and for methods 6 (Fig. 4a) and 7 (Fig. 4b). Similar load variation patterns were obtained for nitrate and ammonium. Results for method M6 were presented because they are representative of methods M1 to M6, whereas method M7 resulted in a load variation pattern different from the six other methods. Examination of Fig. 4 clearly shows the presence of a vertical gap in the load estimation for the linear interpolation method (Fig. 4b, M7), a gap that is not present in the case of the other methods. With method M7, the vertical gap observed in the case of total nitrogen load (Fig. 4b) is not present for sediment.

Figure 4 shows an increase in the range of load estimates with a decrease in the sampling frequency, a phenomenon partly due to the larger number of realizations for low sampling frequency. The use of a fixed or random sampling scheme does not seem to affect the range of calculated load. For a given scheme and frequency, the variation of calculated load is important. For example, Fig. 4a shows that for the fixed sampling scheme, the load variation is small for the daily frequency (2.08 to 3.24 kg/h) and is maximal for the monthly frequency (0.35 to 20.65 kg/h).

The vertical gap in load estimates (Fig. 4b) is present for all nitrogen constituents and all but the daily frequency. The presence of low and high values evokes a hysteretic behaviour. Hysteresis of concentration-flow relationships has been reported for many contaminants (Tracey et al. 2003; House and Warwick 1998). Hysteresis, which is observed for the nitrogen components but not for sediment (Fig. 5), explains the separation of load estimates into two groups for the linear interpolation method. Contrary to the other methods, linear interpolation fills in the missing concentrations in order to obtain a complete set of hourly flow-concentration values. The large number of flow-concentration values reveals the hysteretic behavior of nitrogen, for which every flow value has two corresponding concentrations and, therefore, two calculated load values, a high one and a low one.

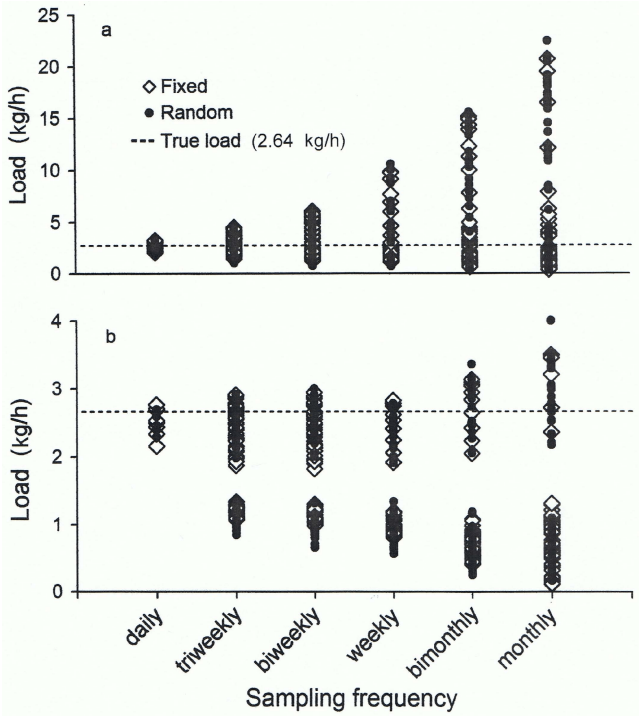

Fig. 4. Variation of load estimation of total nitrogen for all realizations and methods M6 (a) and M7 (b).

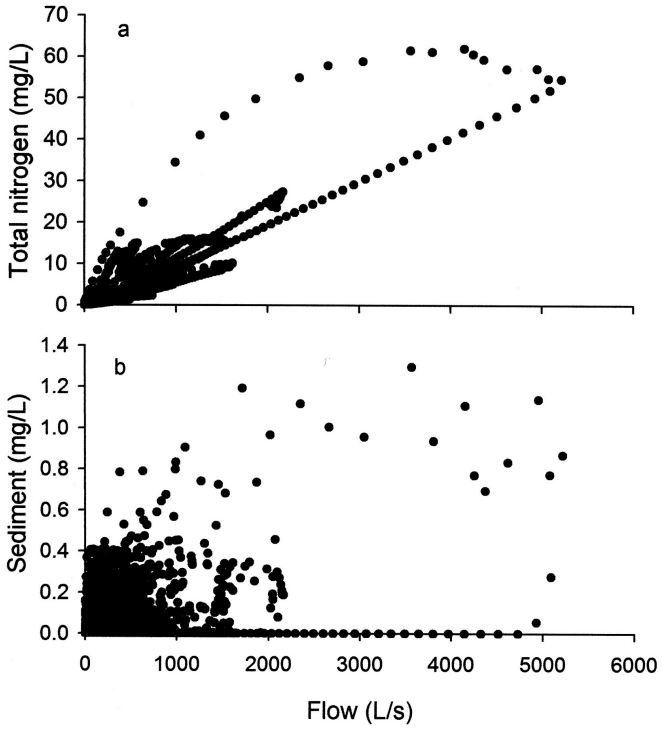

Fig. 5. Hourly flow versus total nitrogen (a) and sediment (b).

CONCLUSIONS

Two years of hourly hydrological and water quality data were obtained from the HSPF model calibrated for a 5.3 km^2 agricultural watershed in Québec. Generated hourly values of flow, sediment, total nitrogen, nitrate, and ammonium were used to evaluate the effect of water quality sampling frequency, sampling scheme, and calculation method on load estimation. Results from this study are summarized as:

1. For all water quality parameters, the quality of load estimation, quantified in terms of the root mean square error (RMSE) between the calculated and true loads, decreased with the sampling frequency.
2. Fixed sampling schemes (e.g. same day of the month for each month) always resulted in lower RMSE values than random schemes (e.g. any day of the month). This better performance of the fixed sampling scheme is attributed to the autocorrelation in time for all water quality parameters. The sediment autocorrellogram showed a 24-hour periodicity that is explained by snowmelt and more frequent evening and night rainfalls.
3. Of the seven load calculation methods studied, linear interpolation, which is used to obtain an estimate of concentration for each available hourly flow, systematically yielded the lowest RMSE value. However, ratio estimator methods did not fare well, ranking only 5 and 6 out of the seven methods tested.
4. All load evaluations generally resulted in an underestimation of the true load. This underestimation is consistent with the percentage of the two-year period with an hourly load below the true load, which ranges from 89% for ammonium to 93% for sediments.

From this study, it results that the most precise and accurate load estimate combines linear interpolation with a fixed sampling scheme. Also, better results were obtained when increasing the sampling frequency up to three samples per week, a scheme for which results were not significantly different from those using a daily sampling. These conclusions apply to the relatively small agricultural watershed used in this study.

REFERENCES

Baker, D.B., R.P. Richards, T.T. Loftus and J.W. Kramer. 2004. A new flashiness index: Characteristics and applications to Midwestern rivers and streams. *Journal of the American Water Resources Association* 40(2): 503-522.

Beale, E.M.L. 1962. Some uses of computers in operational research. *Industrielle Organisation* 31(1): 27-28.

Bergman, M.J., W. Green and L.J. Donnangelo. 2002. Calibration of storm loads in the south Prong watershed, Florida, using BASINS/HSPF. *Journal of the American Water Resources Association* 38(5): 1423-1436.

Bernier, C. and J. Gallichand. 1999. Modélisation hydrologique de la qualité de l'eau d'un bassin versant agricole au Québec. *Canadian Agricultural Engineering* 41: 213-220.

Bicknell, B.R., J.C. Imhoff, J.L. Kittle, A.S. Donigian and R.C. Johanson. 1993. Hydrological simulation program Fortran (HSPF). User's manual 10.1. United States Environmental Protection Agency (USEPA), Athens, GA.

Chambers, P.A., M. Guy, G. Grove, R. Kent, E. Roberts and C. Gagnon. 2002. *Nutrient Losses from Agriculture: Effects on Canadian Surface and Ground Waters*. Publication No. 273: 379-384. Oxfordshire, UK: International Association of Hydrological Sciences.

Chew, C.Y., L.W. Moore and R.H. Smith. 1991. Hydrological simulation of Tennessee's North Reelfoot Creek watershed. *Journal of the Water Pollution Control Federation* 63(1): 10-16.

Chokmani, K. and J. Gallichand. 1997. Utilisation d'indices pour évaluer le potentiel de pollution diffuse sur deux bassins versants agricoles. *Canadian Agricultural Engineering* 39(2): 113-122.

Cochran, W.G. 1977. *Sampling Techniques*, 3rd edition. New York, NY: John Wiley and Sons.

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

Cohn, T.A., D.L. Caulder, E.J. Gilory, L.D. Zinjuk and R.M. Summers. 1992. The validity of sample statistical model for estimating fluvial constituent loads: An empirical study involving nutrient loads entering Chesapeake Bay. *Water Resources Research* 28(9): 2353-2363.

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

Donigian, A.S. and W.C. Hubert. 1991. Modeling of non-point source water quality in urban and non-urban areas. EPA-600/3-19-039. Athens, GA: United States Environmental Protection Agency.

Donigian, A.S., R.V. Chinnaswamy and R.F. Carsel. 1995. Watershed modeling of agrichemicals in Walnut Creek, Iowa. In *Water Quality Modeling,* ed. C. Heatwole, 192-201. ASAE Publication 5-95. St. Joseph, MI: ASABE.

Efron, B. and R.J. Tibshirani. 1993. *An Introduction to the Bootstrap Monograph on Statistics and Applied Probability 57*. Boca Raton. FL: Chapman & Hall / CRC.

Environnement Canada. 1993. Normales climatiques au Canada, 1961-1990. EN56-61/5-1993. Environnement Canada, Ottawa, ON.

Ferguson, R.I. 1987. Accuracy and precision of methods for estimating river loads. *Earth Surface Processes and Landforms* 12: 95-104.

Fogle, A.W., J.L. Taraba and J.S. Dinger. 2003. Mass load estimation errors utilizing grab sampling strategies in a karst watershed. *Journal of the American Water Resources Association* 39(6): 1361-1372.

Gallichand, J., E. Aubin, P. Baril and G. Debailleul. 1998. Water quality improvement at the watershed scale in an animal production area. *Canadian Agricultural Engineering* 40(2): 67-77.

Gangbazo, G., D. Cluis and C. Bernard. 1994. Contrôle de la pollution diffuse agricole à l'échelle du bassin versant. *Sciences et techniques de l'eau* 27(2): 33-39.

Guo, Y., M. Markus and M. Demmissie. 2002. Uncertainty of nitrate-N load computations for agricultural watersheds. *Water Resources Research* 38(10): 1 - 12.

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

Harmel, R.D., K.W. King and R.M. Slade. 2003. Automated storm water sampling on small catchments areas. *Applied Engineering in Agriculture* 19(6): 667-674.

Hébert, S. and S. Légaré. 2000. Suivi de la qualité des rivières et petits cours d'eau. Envirodoq No. ENV-2001-0141, rapport No. QE-123. Québec, Direction du suivi de l'état de l'environnement, ministère de l'environnement.

Hébert, S. and M. Ouellet. 2005. Le Réseau-rivières ou le suivi de la qualité de l'eau des rivières du Québec. Envirodoq No. ENV/2005/0263, collection No. QE/196. Québec, ministère du Développement durable, de l'Environnement et des Parcs, Direction du suivi de l'état de l'environnement.

House, W.A. and M.S. Warwick. 1998. Intensive measurement of nutrient dynamics in the River Swale. *The Science of Total Environment* 210/211: 111-137.

Kauark Leite, L.A. 1990. Réflexions sur l'utilité des modèles mathématiques dans la gestion de la pollution diffuse d'origine agricole. Presented thesis. Paris, France: L'École Nationale des ponts et Chaussées, CERGRENE.

King, K.W. and R.D. Harmel. 2003. Considerations in selecting a water quality sampling strategy. *Transactions of the ASAE* 46(1): 63-73.

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(11): 1483-1501.

Laroche, A.M., J. Gallichand, R. Lagacé and A. Pesant. 1996. Simulating atrazine transport with HSPF in an agricultural watershed. *Journal of Environmental Engineering* 122(7): 622-630.

Legeas, M. and L. Iachkine. 1992. Etude des flux d'azote à l'échelle d'un petit bassin versant : Comparaison des différentes méthodes de calcul. *Techniques Sciences Méthodes (T.S.M.-L'EAU)* 5: 239-244.

Littlewood, I.G. 1992. Estimating contaminant loads in rivers: A review. Report 117. Institute of Hydrology, Wallingford, UK.

Littlewood, I.G. 1995. Hydrological regimes, sampling strategies, and assessment of errors in mass load estimates for United Kingdom rivers. *Environment International* 21(2): 211-220.

Martin, G.R., J.L. Smoot and K.D. White. 1992. A comparison of surface-grab and cross sectionally integrated stream-water-quality sampling methods. *Water Environment Research* 64(7): 866-876.

Miller, P.S., B.A. Engel and R.H. Mohtar. 2000. Sampling theory and mass load estimation from watershed water quality data. ASAE Paper No. 003050. St. Joseph, MI: ASABE.

Moatar, F. and M. Meybeck. 2005. Compared performance of different algorithms for estimating annual nutrient loads flowd by the eutrophic River Loire. *Hydrological Processes* 19: 429-444.

Painchaud, J. 1996. La qualité de l'eau au Québec : État et tendances. Dans *Stratégie de gestion : vers une vision commune. 2e colloque sur la gestion de l'eau en milieu rural*, 37-50. Conseil des Production Végétales du Québec (CPVQ), Québec, QC.

Peart, M.R. 1995. *Sediment and Water Quality in River Catchments: Fingerprinting Sediment Sources, an Example from Hong Kong*. New York, NY: John Wiley and Sons.

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(7): 1035-1050.

Preston, S.D., V.J. Bierman Jr. and S.E. Sillman. 1989. An evaluation of methods for the estimation of tributary mass load. *Water Resources Research* 25(6): 1379-1389.

Rekolainen, S., M. Posch, J. Kämäri and P. Ekholm. 1991. Evaluation of the accuracy and precision of annual phosphorus load estimates from two agricultural basins in Finland. *Journal of Hydrology* 128: 237-255.

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(12): 3747-3759.

Robinson, R.B., M.S. Wood, J.L. Smoot and S.E. Moore. 2004. Parametric modeling of water quality and sampling strategy in a high-altitude appalachian stream. *Journal of Hydrology* 287: 62-73.

Sivakumar, B. and W.W. Wallender. 2004. Deriving high-resolution sediment load data using a nonlinear deterministic approach. *Water Resources Research* 40(5): W054031-W054038.

Stålnacke, P., A. Grimvall, K. Sundblad and A. Tonderski. 1999. Estimation of riverine loads of nitrogen and phosphorus to the Baltic sea, 1970-1993. *Environmental Monitoring and Assessment* 58: 173-200.

Stone, K.C., P.G. Hunt, J.M. Novak, M.H. Johnson and D.W. Watts. 2000. Flow-proportional, time-composited, and grab sample estimation of nitrogen export from an eastern costal watershed. *Transactions of the ASAE* 43(2): 281-290.

Sundborg, A. and A. Rapp. 1986. Erosion and sedimentation by water: Problems and prospects. *Ambio* 15(4): 215-225.

Swistock, B.R., P.J. Edwards, F. Wood and D.R. Dewalle. 1997. Comparison of methods for calculating annual solute exports from six forested Appalachian watersheds. *Hydrological Processes* 11: 655-669.

Tate, K.W., R.A. Dahlgren, M.J. Singer, B. Allen-Diaz and E.R. Atwill. 1999. Timing, frequency of sampling affects accuracy of water-quality monitoring. *California Agriculture* 53(6): 44-48.

Tebbutt, T.H.Y. 1998. *Principles of Water Quality Control*, 5th edition. Oxford, UK: Butterworth-Heinemann.

Toma, A., P. Hubert and M. Meybeck. 1993. Estimation of rivers dissolved and suspended loads interpolating methods – Accuracy and precision. In *First International Conference on Water Pollution: Modeling, Measuring and Prediction.,* 321-328. Ashurst, Southampton, UK: Computational Mechanics Publishing.

Tracey, H.G., A.R. Young, M.G.R. Holmes, G.H. Old, N. Hewitt, G.J.L. Leeks, J.C. Packman and B.P.G. Smith. 2003. The temporal and spatial variability of sediment transport and yields within the Bradford Beck catchment, West Yorkshire. *The Science of Total Environment* 314/316: 475-494.

USEPA. 2003. *National Management Measures to Control Nonpoint Source Pollution from Agriculture*. EPA-841-B-03004. United States Environmental Protection Agency, The National Agriculture Compliance Assistance Center, Kansas City, KS.

Vandenberghe, V., A. van Griensven and W. Bauwens. 2002. Detection of the most optimal measuring points for water quality variables: Application to the river water quality model of the River Dender in ESWAT. *Water Science and Technology* 46(3): 1-7.

Walker, W.W. 1999. Water operations technical support program, simplified procedures for eutrophication assessment and prediction: User manual. Instruction Report W-96-2. US Army Corps of Engineers, Waterways Experiment Station (WES), Vicksburg, MS.

Wang, X., J.R. Frankenberger and E.J. Kladivko. 2003. Estimating nitrate-n losses from subsurface drains using variable water sampling frequencies. *Transactions of the ASAE* 46(4): 1033-1040.

Whelan, M.J., C. Gandolft and G.B. Bischetti. 1999. A simple stochastic model of point source solute transport in rivers based on gauging station data with implications for sampling requirements. *Water Research* 33(14): 3171-3181.

White, E.T. 1999. Automated water quality monitoring. Aquatic inventory task force field manual version 1.0. British Columbia, Ministry of Environment Lands and Parks, Water Management Branch, Resources Inventory Committee, Victoria, BC.

Williams, M., C. Hopkinson, E. Rastetter and J. Vallino. 2004. N budgets and aquatic uptake in the Ipswich River basin, northeastern Massachusetts. *Water Resources Research* 40: W11201 1-12.