HYDROLOGICAL PROCESSES
Hydrol. Process. **20**, 395–410 (2006)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hyp.6057

Accuracy and precision of different sampling strategies and flux integration methods for runoff water: comparisons based on measurements of the electrical conductivity

Patrick Schleppi,* Peter A. Waldner and Bruno Fritschi
Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), CH-8903 Birmensdorf, Switzerland

Abstract:

Because of their fast response to hydrological events, small catchments show strong quantitative and qualitative variations in their water runoff. Fluxes of solutes or suspended material can be estimated from water samples only if an appropriate sampling scheme is used. We used continuous in-stream measurements of the electrical conductivity of the runoff in a small subalpine catchment (64 ha) in central Switzerland and in a very small (0·16 ha) subcatchment. Different sampling and flux integration methods were simulated for weekly water analyses. Fluxes calculated directly from grab samples are strongly biased towards high conductivities observed at low discharges. Several regressions and weighted averages have been proposed to correct for this bias. Their accuracy and precision are better, but none of these integration methods gives a consistently low bias and a low residual error. Different methods of peak sampling were also tested. Like regressions, they produce important residual errors and their bias is variable. This variability (both between methods and between catchments) does not allow one to tell *a priori* which sampling scheme and integration method would be more accurate. Only discharge-proportional sampling methods were found to give essentially unbiased flux estimates. Programmed samplers with a fraction collector allow for a proportional pooling and are appropriate for short-term studies. For long-term monitoring or experiments, sampling at a frequency proportional to the discharge appears to be the best way to obtain accurate and precise flux estimates. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS water chemistry; runoff; sampling schemes; accuracy; catchment hydrology; electrical conductivity; solute fluxes

INTRODUCTION

Headwater catchments are advantageous research objects because they integrate processes over a land area towards a single point, the water outlet. This feature allows monitoring of short- and long-term ecological and geochemical changes in whole ecosystems (Likens and Bormann, 1995; Church, 1997). Small or very small catchments of less than 1 km^2 or even less than 1 ha are especially adapted to experimental research because their size allows manipulations and comparisons. Homogeneity of soil and vegetation, along with the possibility of experimental replications, are further potential advantages.

In large catchments, variations in discharge and water quality are smoothed and appear mainly as seasonality (e.g. Davis, 1986). In small catchments, however, water discharge and water quality fluctuate more (Keller *et al.*, 1989). Short-term variations are not smoothed because of the short residence times and because of limited effects of pooling between subcatchments. Especially in small catchments, measuring fluxes of solutes or sediments therefore requires (1) a discharge gauging with a broad measuring range, (2) an adapted water sampling scheme, and (3) a correct numerical integration of concentrations and discharges over time. In

* Correspondence to: Patrick Schleppi, WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland. E-mail: schleppi@wsl.ch

Received 27 January 2004
Accepted 9 February 2005

designing a sampling scheme, the accuracy and precision of the measurements have to be optimized by taking account of the costs of (1) field instrumentation, (2) sampling and maintenance, including travelling to the field, and (3) laboratory work and chemical analyses.

Gauging stations with a continuous (or at least very frequent) measurement of the discharge are relatively easy to design. Online water analyses, on the contrary, are only possible for a limited number of parameters, with high instrumentation costs (Pacini *et al.*, 1997) and not always with the desired analytical accuracy (Ranalli, 1998). Laboratory sample analyses are thus usually done at a much lower frequency than the discharge measurements. Common sampling schemes include manual grab sampling, programmed automatic sampling, automatic event (peak) sampling and flow-proportional sampling. Several methods are used to integrate the concentration and discharge data into fluxes, including graphical expert judgement, interpolation, regressions, and other estimators described below. Flux estimates based on grab sampling and simple multiplications (Equation (1)) generally are recognized as biased:

$$F = Qc\Delta t \tag{1}$$

where F is the flux, Q is the discharge, c is concentration and Δt is the time interval. The bias is positive if the concentration is negatively correlated to the discharge, and vice versa (Schwartz and Naiman, 1999).

Several studies have been published comparing different sampling schemes and/or integration methods, either in a mathematical approach (e.g. Cohn, 1995; Schwartz and Naiman, 1999) or based on empirical measurements (Johnson, 1979; Walling and Webb, 1985; Dann *et al.*, 1986; Preston *et al.*, 1992; Thomas and Lewis, 1993; Swistock *et al.*, 1997; Bukaveckas *et al.*, 1998; Line *et al.*, 1998; Robertson and Roerish, 1999; Stone *et al.*, 2000; Coats *et al.*, 2002; Cooper and Watts, 2002). Comparisons usually consider both the accuracy (low systematic error, i.e. low bias) and the precision (low residual error) of the estimates. As a reference, some authors use the method that they believe to be the best; in several cases, however, such a reference method has either no estimate of its own bias, or is even described as biased by other authors. A more objective approach is to make frequent analyses and to use the full set of results as a reference for comparison with different methods based on subsamples out of this population. In the best cases (Preston *et al.*, 1992; Robertson and Roerish, 1999), daily analyses were available as a reference. Even in large catchments, Preston *et al.* (1992) found that most load estimates are biased if concentrations are a function of the discharge and if the latter shows a high variability. Whereas almost all methodological comparisons were made on catchments larger than 1 km^2, Cohn (1995) pointed out that short discharge peaks must be taken into account when designing a sampling scheme for smaller catchments. Since precipitation events immediately affect their runoff, samplings at intervals shorter than a day appear to be necessary. Relationships between discharge and solute concentrations can vary within a single event, depending mainly on the contributions of new and old water. Fertilizers and atmospheric inputs, processes in the snowpack, exchange, transformation, flushing and uptake in the soil can further complicate these relationships, depending on the solute and flow regime considered. An ideal sampling scheme should, therefore, be robust enough not to be biased by such processes.

From the published comparisons, 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. In the present paper, we take advantage of continuous in-stream measurements from both a small subalpine catchment and a very small subcatchment to simulate different methods and to compare them with the complete integration of the available data.

MATERIAL AND METHODS

Catchments and instrumentation

The Erlenbach catchment (0·64 km^2, in the Alptal valley, central Switzerland) was chosen for this study, along with an experimental subcatchment that is 400 times smaller (Figure 1). The geological parent material of the valley is Flysch, a formation typical of the northern edge of the Alps consisting of alternating calcareous sandstones with argillite and bentonite schist's. Soils are umbric Gleysols. Because of the very low permeability of these materials, there is practically no deep infiltration of water. The Erlenbach catchment lies at an altitude between 1100 and 1600 m, has a westerly aspect and an average slope of 20%. Up to 40% of the catchment is covered by forests and 60% by wet grassland. The climate is cool and wet: 6 °C average air temperature and 2200 mm precipitation per year (altitude of the weather station: 1200 m) (Burch, 1994). The experimental subcatchment (1600 m^2) is located near the weather station and covered by a forest of Norway spruce (*Picea abies*) with some silver fir (*Abies alba*). It was chosen as a control plot for the research project 'Nitrex' (nitrogen saturation experiments; Schleppi *et al.*, 1998) and, therefore, is referred to hereafter as the Nitrex subcatchment.

The Erlenbach catchment and the Nitrex subcatchment are each equipped with a V-notch weir to measure their water discharge (Burch, 1994; Schleppi *et al.*, 1998). For 2 years, starting from June 1995, the electrical conductivity of the runoff was measured in both catchments.

Electrical conductivity and water chemistry

The electrical conductivity of water is a function of its ion content. This function is linear at low concentrations, but a non-linear correction has to be introduced at higher ionic strengths (Laxen, 1977). Applied to our chemical analyses, this correction was found to be between −0·1 and −13%. Therefore, when water samples are pooled, ion concentrations add, but conductivities not exactly. However, this non-additivity was always small (<0·5%), even when mixing samples with extremely contrasting ionic strengths. Because of the high conductance of H^+ (H_3O^+) ions, acid–base reactions can also affect the conductivity of water samples. In the runoff of the catchments studied, pH values are neutral (between 7 and 8) and protons, therefore, contribute negligibly (<0·05%) to the electrical conductivity. In conclusion, the electrical conductivity of our runoff waters behaves linearly upon mixing. Therefore, it is possible to integrate it over time in the same way as concentrations are integrated to fluxes.

Simulation of sampling strategies

Compared with chemical analyses, electrical conductivity is very easy to measure and can be monitored with a high temporal resolution. The data available for the present study were averaged and recorded every 10 min. This large number of data (approximately 100 000 measurements over 2 years) allowed us to simulate different sampling strategies and to compare their ability to produce representative weekly samples. The four strategies considered were fixed-time, volume-proportional, frequency-proportional and peak samplings. Results from fixed-time (grab) sampling were further used to test different integration methods. Compared with these simulations, solutes with different flow dependencies would certainly give other sampling errors. Knowing this, our main goal here is to find general rules and to rank methods. A comparison between different solutes will be part of another contribution based on a real implementation of two selected sampling schemes (fixed time and frequency proportional, Schleppi *et al.*, 2005).

Fixed-time sampling (FT). This first approach consists of taking individual samples with a defined frequency and pooling them into a composite sample. It is the easiest sampling strategy, as it requires only manual sampling or, at higher frequencies, a programmable sampler. Therefore, it is a widely used method. To simulate a single, fixed-time sample, we took the value of the electrical conductivity in the middle of the week to represent a grab sample. Each measurement is then multiplied by the weekly water discharge to give a flux. This corresponds to method number 3 by Walling and Webb (1985). We also calculated average values of the conductivity measured once a day, every 8 h or every hour. These fixed-time methods were denoted FT1, FT3 and FT24 according to the number of samples per day. In all cases, the water discharge was calculated from the complete dataset, with the available temporal resolution of 10 min.

Volume-proportional sampling (VPi, VPa). Here also, the individual samples are taken based on a fixed schedule. They are, however, pooled in proportions based on the corresponding discharges. Practically, it is possible to use a water sampler with a fraction collector, the individual samples later being combined manually in the correct proportions. Some collectors allow for a sampling volume controlled by a measurement of the discharge and may thus be used to take a proportional bulk sample. Again, the simulation was done for either 1, 3 or 24 individual samples per day. These volume-proportional methods were named VPi1, VPi3 and VPi24, the letter 'i' denoting that the instantaneous discharges were used for pooling the samples. A variation of this method is obtained by mixing proportions based on the discharge averaged over the periods of time around each sampling (strategies VPa1, VPa3 and VPa24, possible only with a fraction collector).

Frequency-proportional sampling (FP). The third strategy supposes that the continuous measurement of the discharge is used to drive the sampling. It requires an automatic sampler coupled to the gauging station: every time a defined volume of water has flowed by, a signal (impulse) is given to actuate the sampler. This is made possible by using either an electro-mechanical device, as described by Fredriksen (1969), or, nowadays more precisely and easily, with an electronic unit calculating the discharge and integrating it online. In the simulation, the value of the conductivity was taken every time the cumulated discharge reached 1, 0·1 or 0·01 cm. The number of samples per centimetre was used to denote the methods: FP1, FP10 and FP100. During periods of low discharge, it is possible that low sampling frequencies do not generate any sample (no impulse). In that case, a grab sample has to be taken at the end of the sampling period.

Peak sampling (P). In as much as most of the water discharge occurs during runoff peaks, it may be appropriate to concentrate sampling efforts on them, especially in small catchments. Therefore, we simulated samples taken at fixed times, but only if the discharge was above a baseline value. This method requires a time-programmable sampler coupled to a switch responding to the chosen minimum discharge. We chose three different values for the baseline: 2 µm min^{-1}, 4 µm min^{-1} and 8 µm min^{-1} for the methods P2, P4 and P8 respectively. These discharges are near or somewhat above the maximum baseflow obtained from the procedure of Jordan *et al*. (1997) (Erlenbach: 2 µm min^{-1}; Nitrex: 1 µm min^{-1}).

Two other methods were tested with samples simulated at single stages (5, 10, 20 and 40 µm min^{-1}) of discharge peaks, either during the ascending limb (PS1) or during both the ascending and descending limbs (PS2). Method PS2 may be better adapted for solutes tending to show hysteresis, e.g. with concentration changes preceding discharge peaks. Finally, sampling was also simulated at the maximum of discharge peaks (P), a peak being recognized as such when the discharge increased at a rate above 0·02 µm min^{-2} (Erlenbach) or 0·03 µm min^{-2} (Nitrex). These values were chosen to identify short-lived peaks from short rainfall, as well as slow-rising peaks from daily snowmelt, without responding to measurement noise. A variant (PP) of this peak sampling consisted of combining the single samples proportionally to the corresponding maximum discharges. As in weeks without impulses for FP, grab samples were simulated at the end of weeks without identified flow peaks. These sampling methods require both an automatic sampler and an electronic device to recognize the discharge peaks in real time. For PP, a fraction collector is also necessary.

Test of integration methods

Because of the bias generally associated with the simple multiplication of fixed-time concentrations and discharges (Equation (1)), several other integration methods have been proposed. Their common goal is to improve the accuracy of flux estimates without having to implement a more complicated sampling scheme. These integration methods can be classified into two groups: those that estimate a concentration at the same frequency as the discharge measurements (interpolation and regression methods), and those that give only flux estimates over periods covering several samplings (average estimates).

Regression methods (R). Concentrations of solutes are often correlated (positively or negatively) to discharge. To account for this effect, it is common practice to use a regression predicting concentrations from discharge measurements. Weekly samples (like in FT) were used to calculate regressions of the electrical conductivity; the 104 data points from 2 years of sampling were used. Log–linear regressions (Equation (2)) were calculated according to Swistock *et al.* (1997); like these authors, we considered seasonality effects as harmonic terms, including their interaction with the discharge. In addition, we further tested a log–quadratic and a log–log model (Equations (3) and (4) respectively).

$$L = a + b \log Q + f(t) \tag{2}$$

where

$$f(t) = c_1 \cos(t) + c_2 \sin(t) + c_3 \cos(2t) + c_4 \sin(2t) + c_5 \cos(3t) + c_6 \sin(3t) + c_7 \log Q \cos(t) + c_8 \log Q \sin(t) + c_9 \log Q \cos(2t) + c_{10} \log Q \sin(2t) + c_{11} \log Q \cos(3t) + c_{12} \log Q \sin(3t)$$

and where L is electrical conductivity, Q is water discharge, $t = 2\pi/365$ days of the year, and a, b, and c_1 to c_{12} are regression coefficients.

$$L = a + b_1 \log Q + b_2(\log Q)^2 + f(t) \tag{3}$$

$$\log L = a + b \log Q + f(t) \tag{4}$$

Average estimates. Giving more importance to concentrations measured at high water discharge compared with those measured at low flow can reduce the bias of flux estimates. We therefore tested a mean weighted by the discharge at sampling time (Mi), which corresponds to the often-used method 5 of Walling and Webb (1985). We chose to apply it to eight weekly grab samples at a time, which gave 13 estimates for the 2 years considered, allowing us to calculate standard deviations.

Beale's ratio estimator (BRE) was further calculated according to Cohn (1995). As proposed by Preston *et al.* (1992), we also tried to stratify the data according to the discharge prior to the calculations, again with 13 groups of eight cases (BREs).

Evaluation of the simulated strategies

All the strategies considered were compared with an ideal sampling represented by the available 10 min resolution. We consider this frequency sufficient to describe the quantitative and qualitative dynamics of the runoff. Sampling and integration methods were compared with this ideal sampling; differences were interpreted as consisting of a bias (systematic error) and a residual (random) error. As a first step, we compared the strategies at a medium sampling frequency, which is the most likely to be used practically; different frequencies were also compared as a second step.

RESULTS

Discharge and electrical conductivity

As expected for a steep mountain stream, the discharge of the Erlenbach shows a broad variation. In the Nitrex subcatchment, the range was even broader (Figure 2). The standard deviations of the logarithms (base 10) were 0·61 and 0·93, corresponding to geometrical standard deviations of 4·1 and 8·5 respectively. During the period of study, the Erlenbach was dry for 4 days and the Nitrex subcatchment for 62 days. The highest discharges were recorded during a thunderstorm on 14 July 1995, when 36 mm of rain fell within 20 min. The probability of such a rain intensity occurring was estimated to be 1 in 30 years. A baseflow index

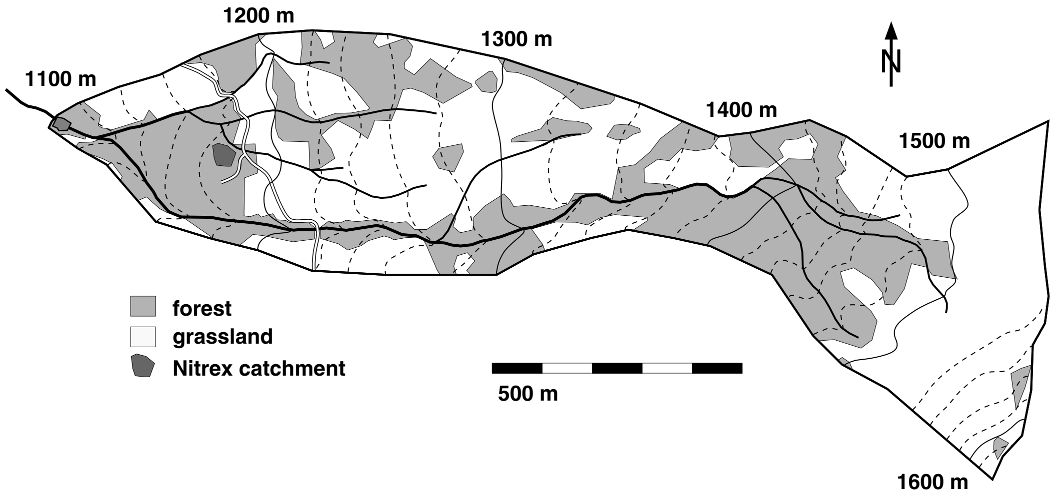

Figure 1. Map of the Erlenbach catchment (Alptal, central Switzerland) including the Nitrex subcatchment

was calculated according to Jordan *et al*. (1997) and gave 19% for the Erlenbach and 4·5% for the Nitrex subcatchment. Most of the discharge thus occurred as peak discharge.

The electrical conductivity was markedly higher in the Erlenbach runoff (weighted average 161 μS cm^{-1}) than in the Nitrex subcatchment (51·5 μS cm^{-1}). This difference corresponds to higher calcium carbonate concentrations in the Erlenbach and is likely related to longer residence times of the water in deeper soil or subsoil layers (Schleppi *et al*., 1998). In both catchments, there was a strong negative correlation with the discharge, indicating a net dilution effect by precipitation water (Figure 2). In the Erlenbach, there was some

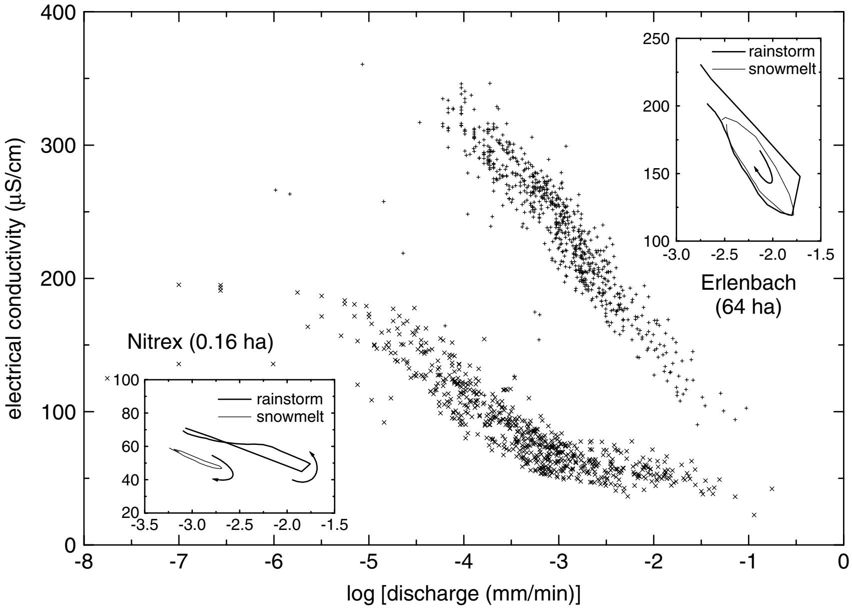

Figure 2. Relationship between the logarithm of the discharge and the water electrical conductivity. Daily values from 2 years in the Erlenbach catchment and the Nitrex subcatchment. Insets show typical curves (hourly values) for a rainstorm and for a day of snowmelt

hysteresis in this relation: changes in conductivity were usually delayed compared with changes in discharge (Figure 2, inset). In the Nitrex subcatchment, there was practically no hysteresis, probably because of the faster hydrological response and, thus, less in-catchment mixing of new and old water.

Differences between sampling strategies

At the middle intensity of three samples per day, the fixed-time sampling FT3 produced a few correct values, but simulated most samples with excess conductivities. When plotted against the ideal proportional sampling (Figure 3), results from FT3 were roughly within a triangle: high conductivities were fairly well simulated, but lower values showed much scatter above the 1 : 1 line. Because low conductivities correspond to high discharges, these overestimations heavily affect flow calculations, resulting in a strong positive bias and a large residual error. Over the period of study, this method overestimated the conductivity flux by +26% for Erlenbach and +36% for Nitrex.

Intensifying the fixed-time sampling from a single grab sample per week (FT) to hourly samples (FT24) only improved the results slightly. Neither in the Erlenbach catchment nor in the Nitrex subcatchment (Table I) did frequent sampling correct for the major fault of this method, i.e. numerous and strong overestimations of the conductivity.

With three samples per day, the volume-proportional sampling VPi3 was better than the fixed-time method, and in many cases was close to the ideal sampling (Figure 4). In other cases, however, it also tended to overestimate the conductivity. This scatter leads to a minor positive bias (Erlenbach: +0·7%; Nitrex: +1·8%) and to some residual error. Using the average discharge over 8 h (instead of the instantaneous measurement) degraded the results slightly and generated more bias (Erlenbach: +3%; Nitrex: +5%).

Increasing the number of samples collected per day clearly improved the estimates, both in terms of bias and residual error, and for both instantaneous (VPi) and average (VPa) discharge references (Table I). Hourly samples pooled proportionally then gave very accurate results (bias within ±0·4%).

With 10 samples collected per centimetre discharge, the frequency-proportional sampling FP10 produced results close to the ideal sampling (Figure 5). Practically no bias occurred. Some scatter could be seen on the Nitrex subcatchment where too low or too high conductivities were simulated during weeks when a grab sample had to be taken because there was no impulse for automatic sampling. Because of their low discharges,

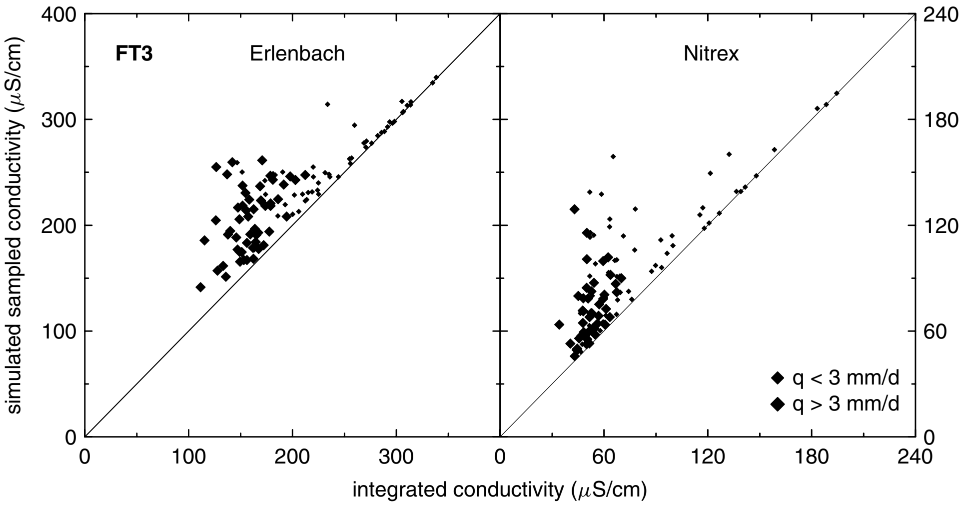

Figure 3. Electrical conductivity obtained from fixed-time sampling (three samples per day, pooled weekly) in comparison with values integrated from measurements every 10 min

Table I. The sampling strategies and their biases (systematic errors) and random (residual) errors; from the simulated weekly fluxes compared with the full integration of electrical conductivities measured every 10 min

Sampling strategy or integration method		Abbreviation	Erlenbach (64 ha)		Nitrex (0·16 ha)	
			Bias (%)	Residual error (%)	Bias (%)	Residual error (%)
Fixed-time sampling	1/week	FT	29·4	44·1	39·2	57·2
	1/day	FT1	29·2	29·1	37·7	41·2
	3/day	FT3	25·8	27·8	36·3	40·2
	24/day	FT24	25·8	28·7	36·3	40·3
Volume-proportional sampling (average Q)	1/day	VPa1	12·6	16·4	13·9	22·5
	3/day	VPa3	2·9	7·4	4·7	11·5
	24/day	VPa24	0·0	1·8	0·2	0·5
Volume-proportional sampling (instantaneous Q)	1/day	VPi1	8·4	17·8	6·8	21·2
	3/day	VPi3	0·7	8·5	1·8	9·6
	24/day	VPi24	−0·3	3·0	−0·4	3·0
Frequency-proportional sampling	1/cm water	FP1	0·3	8·0	0·1	6·0
	10/cm water	FP10	0·0	0·6	0·2	1·2
	100/cm water	FP100	0·0	0·0	0·0	0·1
Peak sampling at maxima	Non-proportional	P	−5·2	13·5	0·1	12·4
	Proportional	PP	−14·0	20·2	−8·0	12·7
Peak sampling (3/h when Q > baseline)	$Q > 2\ \mu m\ min^{-1}$	P2	7·8	20·5	6·1	20·6
	$Q > 4\ \mu m\ min^{-1}$	P4	−0·6	16·9	2·6	15·5
	$Q > 8\ \mu m\ min^{-1}$	P8	−8·5	13·0	0·3	11·6
Peak sampling at single stages	Ascending limb	PS1	4·2	22·1	0·3	20·3
	Both limbs	PS2	0·3	19·9	2·1	19·9
Regressions[a]	Log–linear	R1	0·1	14·0	−28·7	80·8
	Log–quadratic	R2	−1·7	16·5	−2·8	29·5
	Log–log	RL	3·5	10·7	−5·7	22·4
	Log–log corrected[b]	RLc	3·8	10·8	−5·0	22·0
Weighted mean[a]	By instantaneous Q	Mi	14·1	21·4	14·9	21·7
BREs[a]	Unstratified	BRE	11·3	22·8	12·8	22·0
	Discharge-stratified	BRE_s	5·1	14·8	9·0	14·6

[a] These integration methods are based on weekly, fixed-time samples (FT).
[b] Retransformation bias corrected after Ferguson (1986).

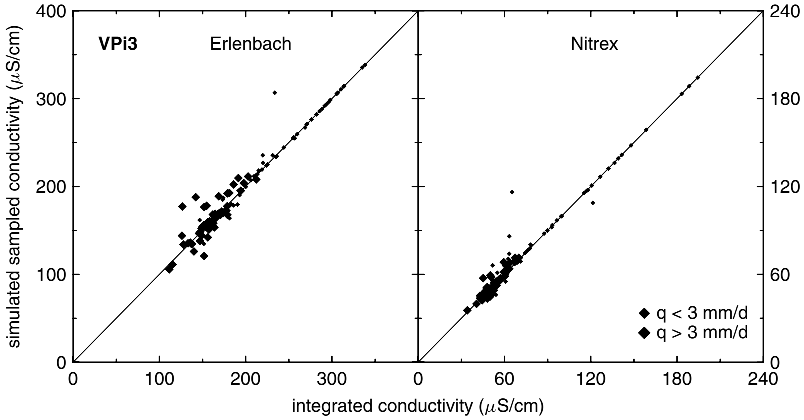

Figure 4. Electrical conductivity obtained from volume-proportional sampling (three samples per day, pooled weekly proportionally to the instantaneous discharge) in comparison with values integrated from measurements every 10 min

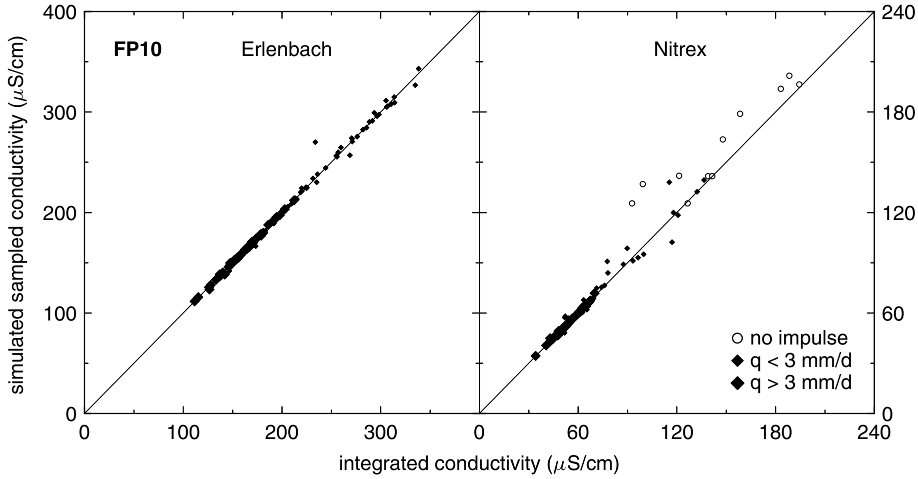

Figure 5. Electrical conductivity obtained from frequency-proportional sampling (10 samples per centimetre discharge, pooled weekly; grab sample when no impulse for a proportional sample) in comparison with values integrated from measurements every 10 min

and despite their high conductivities, these points have little weight in the longer term, and the fluxes are thus estimated very accurately.

Intensifying the frequency-proportional sampling had practically no effect of the bias, which was always very small. The residual error, however, was reduced due to a reduction in the number of cases with no automatically taken sample. At the lower frequency of one sample per centimetre (FP1), there were 14 weeks out of 104 when a grab sample had to be simulated for the Erlenbach. In the Nitrex subcatchment, the corresponding values were 26 with FP1 and 11 with FP10. The highest frequency FP100 always yielded samples.

There were often weeks without peaks for the P and PP methods (Erlenbach: 28 times in 104 weeks; Nitrex: 24 times), and grab samples had to be simulated instead. In these cases, the results were similar to those of the FT method and the conductivity was mostly overestimated (Figure 6). When peaks occurred, the

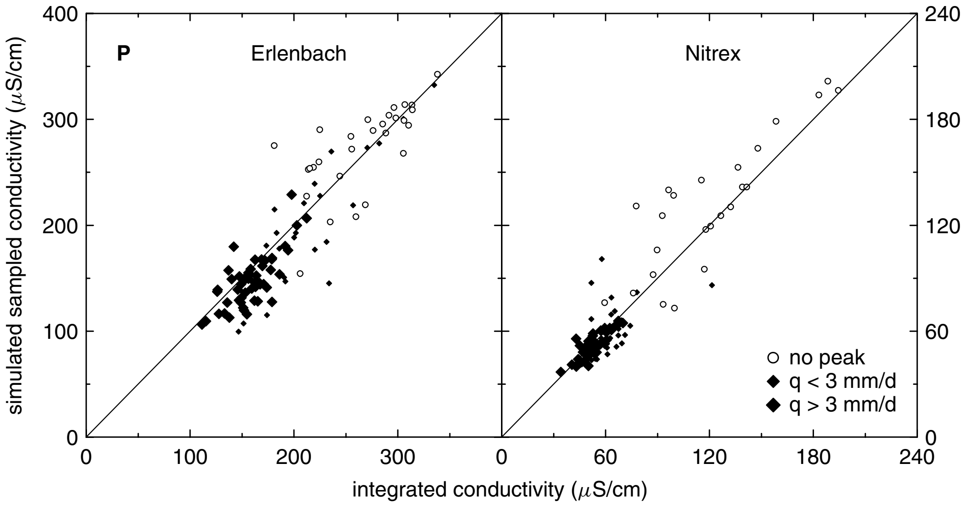

Figure 6. Electrical conductivity obtained from peak sampling (samples taken at discharge peaks, pooled weekly; grab sample when no peak) in comparison with values integrated from measurements every 10 min

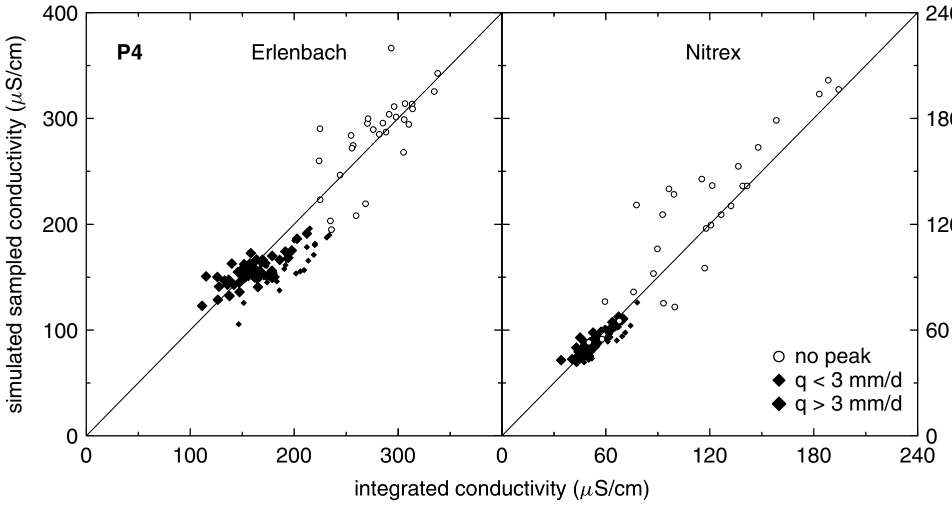

Figure 7. Electrical conductivity obtained from peak sampling (three samples per hour when discharge above a baseline of 4 μm min^{-1}, pooled weekly; grab sample when no peak) in comparison with values integrated from measurements every 10 min

estimates scattered around the 1 : 1 line, more below than above it for the Erlenbach. Because they contribute much to the discharge, these peaks produced an underestimation of the P method in the Erlenbach (−5%), but not in the Nitrex subcatchment. When large peaks were more weighted than small ones (PP), there was more negative systematic error (Erlenbach: −14%; Nitrex: −8%).

The chosen baseline values of the methods P2, P4 and P8 (2 μm min^{-1}, 4 μm min^{-1} and 8 μm min^{-1} respectively) were not reached every week. The number of weeks without sampled peaks increased with these baselines: 21, 29 and 38 weeks for the Erlenbach and 23, 29 and 37 weeks for the Nitrex subcatchment. In weeks with peaks but with a relatively low total discharge, the conductivity was clearly underestimated (Figure 7), but the contrary happened in weeks with a high total discharge. In the Erlenbach, this appeared very clearly as a 'flattening' of the estimates compared with the theoretical values. The bias resulting from this peak sampling decreased with increasing baselines (Table I). In the case of the Erlenbach, nearly unbiased results were obtained for P4, but for the Nitrex subcatchment the higher baseline of P8 was close to the optimum.

Sampling at single peak stages (PS1 and PS2, Figure 8) produced patterns similar to those obtained with P2 to P8. Again, the number of weeks without sampled peaks was important (Erlenbach: 31; Nitrex: 29) and the residual error was relatively large in all cases (Table I). For the Erlenbach, a lower bias was obtained when both the ascending and the descending limbs of hydrograph peaks were sampled (PS2). In the subcatchment, however, it was better to sample only the ascending limb (PS1).

Differences between integration methods

Different results were obtained from the regression methods depending on the catchment and on the model chosen. Log–linear regressions (Equation (2)) from weekly data gave coefficients of determination $r^2 = 0{\cdot}95$ for the Erlenbach and $r^2 = 0{\cdot}88$ for the subcatchment. The model (Figure 9) gave fair estimates when applied to the entire Erlenbach dataset; with some scatter around the ideal line and a slightly convex curve, there was practically no bias. In the Nitrex subcatchment, however, the same model gave severe underestimations in the lower conductivity range, and produced a bias of −29%. The log–linear regression obviously failed to take into account that the conductivity does not decrease so much at very high discharges (Figure 2).

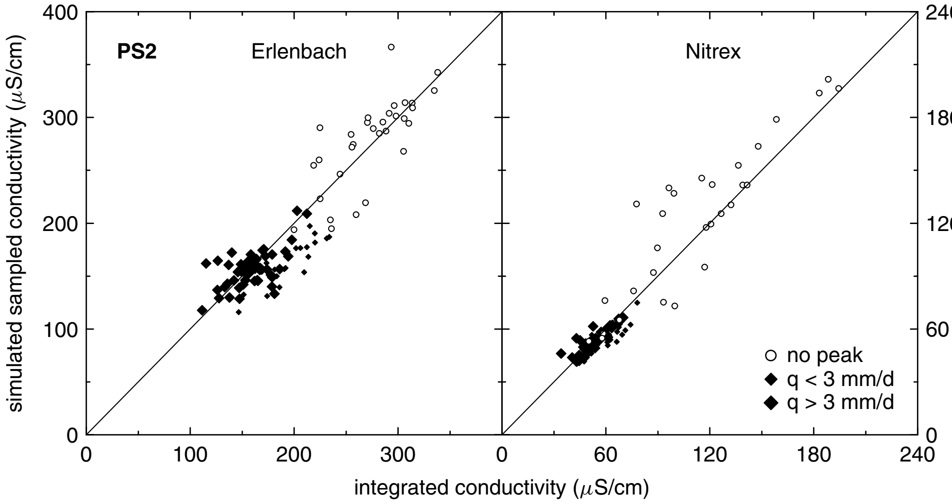

Figure 8. Electrical conductivity obtained from peak sampling at single discharge stages during both ascending and descending limbs (5, 10, 20 and 40 μm min^{-1}, pooled weekly; grab sample when no peak) in comparison with values integrated from measurements every 10 min

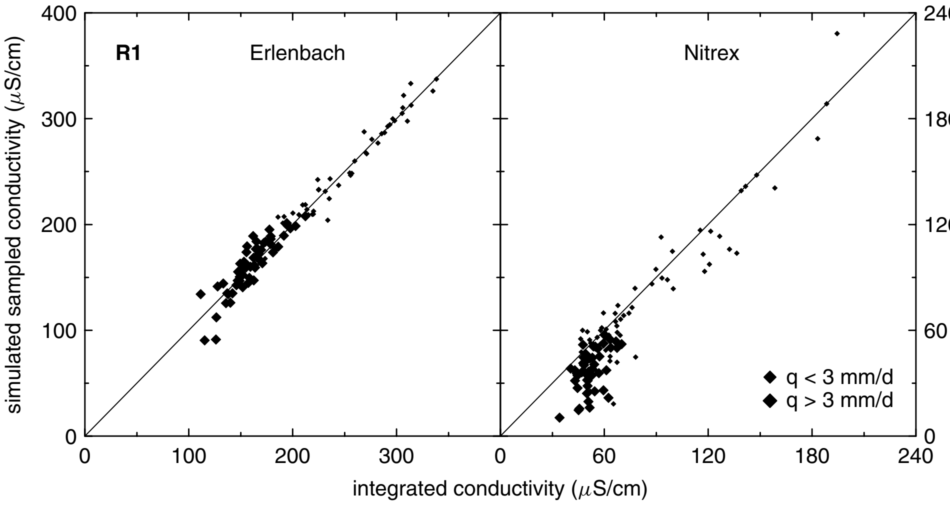

Figure 9. Electrical conductivity obtained from a log–linear regression (calculated from 104 weekly values of conductivity against discharge and seasonality, then applied on 10 min discharge measurements) in comparison with values integrated from measurements every 10 min

In terms of r^2, the log–quadratic model (Equation (3)) was only marginally better (Erlenbach: still 0·95; Nitrex: 0·90). The predictions (Figure 10) were almost the same as from the log–linear model in the case of the Erlenbach. In the subcatchment, however, the gross underestimations were largely corrected, resulting in a much better accuracy (bias −2·8%).

Compared with the other models, log–log regressions (Equation (4)) had a slightly lower r^2 for the Erlenbach (0·93), but a slightly higher r^2 for the subcatchment (0·91). In the main catchment, the model gave the smallest residual error but a small positive bias (3·5%) of the conductivity, because of overestimations in the lower range (concave curve, Figure 11). In the subcatchment, the log–log

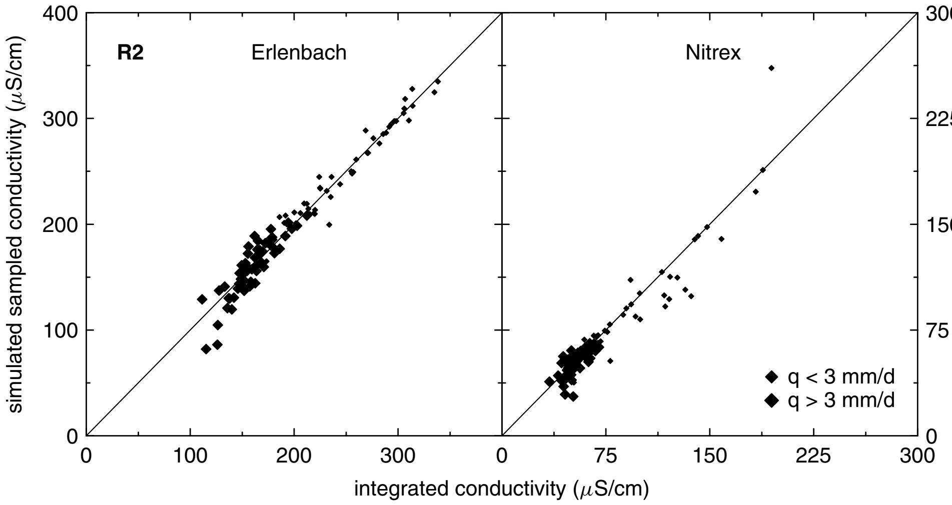

Figure 10. Electrical conductivity obtained from a log–quadratic regression (calculated from 104 weekly values of conductivity against discharge and seasonality, then applied on 10 min discharge measurements) in comparison with values integrated from measurements every 10 min

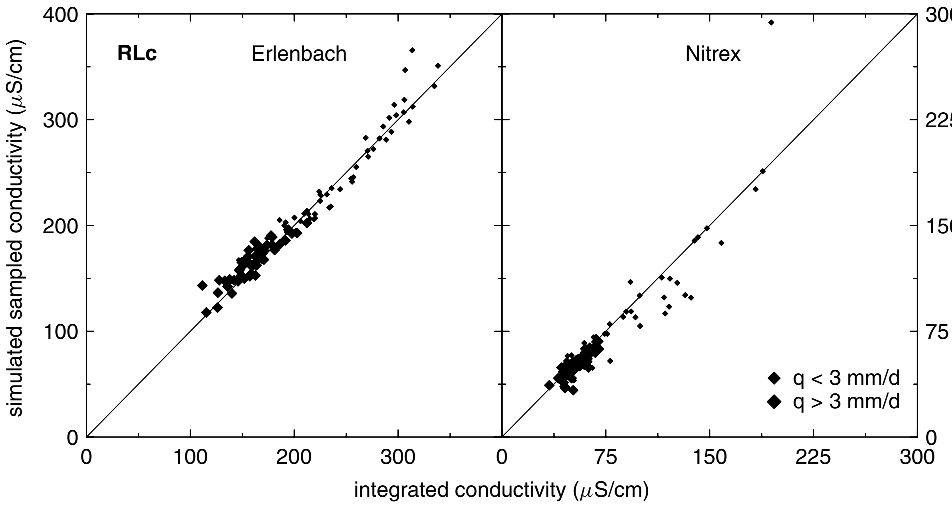

Figure 11. Electrical conductivity obtained from a log–log regression with Ferguson's (1986) correction (calculated from 104 weekly values of conductivity against discharge and seasonality, then applied on 10 min discharge measurements) in comparison with values integrated from measurements every 10 min

model produced an underestimation (−5·7%). The correction factor of Ferguson (1986) was lower than 1% for both catchments and, therefore, had only a very small positive effect on the calculated fluxes.

The mean weighted by instantaneous discharge Mi overestimated the fluxes by 14–15% (Figure 12). This bias was less than half that with FT. Without prior stratification, BREs gave patterns very similar to Mi (details not shown) and overall slightly better results (Table I). After stratification (Figure 12), there was less bias, but still an overestimation in most discharge classes.

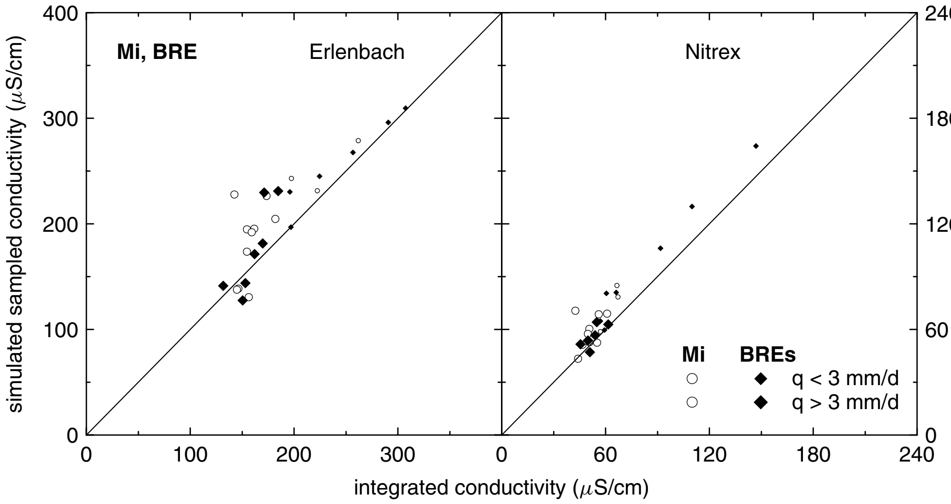

Figure 12. Electrical conductivity obtained as mean weighted by the instantaneous discharge (Mi) as discharge-stratified BRE (BRE_S) from weekly data (13 groups of eight values) in comparison with values integrated from measurements every 10 min

DISCUSSION

Because of their low water retention and fast response to hydrological events, small catchments usually show more variability in their discharge than larger ones. This general rule was verified between our Nitrex and Erlenbach catchments. In both cases, we observed that higher discharges corresponded to lower electrical conductivities. St-Hilaire *et al.* (1998) already noticed that this relationship is strong and that it must be taken into account to calculate fluxes of solutes. Because it fails to do this, a fixed sampling schedule gives too much weight to periods of low flow and thus introduces a bias towards the corresponding concentrations (in our case towards high electrical conductivities). This method may be appropriate when extreme values are of particular interest, as demonstrated by van Sickle *et al.* (1997) for the episodic acidification of streams. Fixed-time samples, however, are fully inappropriate to produce flux estimates of solutes when concentrations and discharge are related and variable.

A volume-proportional sampling gives much better results. It does not require much more instrumentation than a frequent fixed-time sampling would: only a fraction collector, which is an option of most commercially available samplers. Attention must be paid, however, to take individual samples relatively often. If this is not the case, then too many short discharge peaks are missed and flux estimates are correspondingly biased. In our study, three samples per day were necessary to obtain fair estimates. This option can be realized with a 24-bottle fraction collector working over 1 week. Hourly samples give better results, but they require a daily change of sample bottles. Because of the amount of work in collecting and proportionally mixing the individual samples, this method is more appropriate for short-time studies, and less so for long-term monitoring. An advantage is that there is always the possibility of analysing the individual samples separately, especially when an interesting event has occurred. Samplers able to pump variable water volumes into a single bottle do not offer this option. They are further limited by their range of volumes, typically between 20 ml and 10 l. If the entire range must be used, then this results in very large bulked samples. And in catchments like those studied here, even the entire range is too small compared with the variability of the discharge.

A frequency-proportional sampling was shown to give excellent results. Using this sampling method, Moldan and Wright (1998) also observed imprecision at low discharges for nitrate. As in our case, this did not affect the calculation of fluxes over longer time periods. The only parameter to adapt is the discharge per individual sample, which is limited by the time needed for a sample to be pumped, which is typically between

20 s and more than 1 min for commercially available samplers. On the other hand, low frequencies sometimes fail to yield any sample and should, therefore, be avoided. The cost of the instrumentation (impulses to be generated online) is higher than for a fixed-time sampling. In the mid and long term, however, this investment is negligible compared with the possibility of obtaining reliable results with only a few (expensive) chemical analyses of water samples. Frequency-proportional sampling is, therefore, the method of choice, especially for mid- and long-term monitoring. Because the samples are pooled as they are taken, the only disadvantage is that it does not allow detailed analysis of events that *a posteriori* appear particularly interesting.

When samples are taken at the maximum discharge (peak sampling P and PP), online identification of the peaks is required; the costs are thus roughly the same as for frequency-proportional sampling. The results, however, are always less accurate, and this method cannot be recommended. In our simulations, peak sampling above a baseline discharge generates a bias that clearly increases as the chosen baseline value increases. This can be explained by the periods of relatively low discharge, which sometimes fall below and sometimes fall above the baseline; such samples have rather high conductivities (Figure 2). Finding an optimum baseline discharge is thus important; but this is only possible when the joint distribution of discharges and concentrations can be fairly estimated, and thus only after enough analyses. The best baseline discharge for the Erlenbach would be roughly twice the baseflow estimated from hydrographs, but for the Nitrex subcatchment it would be eight times the baseflow. The baseflow thus does not give a good indication of which baseline is to be chosen. As expected, sampling both limbs of peaks (PS2 versus PS1) is more useful when the concentrations show a hysteresis relatively to the discharge (Erlenbach). As was shown by an end-member analysis (Hagedorn *et al.*, 2001), this can be related to variable contributions of different water sources within a discharge event. As a rule, therefore, all parts of discharge peaks should be sampled.

Because peak sampling and fixed-time sampling often give opposite biases, both methods can be combined to produce better results, as shown by Line *et al.* (1998) for phosphorus and nitrate in streams draining pastures. This combination is quite often used (e.g. Swistock *et al.*, 1997). A major drawback, however, is that the optimum relative weight of peak versus grab samples can be known only after enough analyses have been done. Anyway, in our case, some peak sampling methods gave a positive bias, like grab samples. And even if a good accuracy can be achieved, the laws of error propagation predict that a large residual error will remain.

All the regressions and average estimators brought improvements in bias and residual error compared with the underlying grab samples. The smaller bias of Mi compared with FT confirms the results of Walling and Webb (1985) with their corresponding methods 5 and 3 respectively. BRE was better than Mi, both in terms of bias and of residual error. Stratifying the data according to the weekly discharges (BREs) further reduced the errors, as described by Preston *et al.* (1992). This shows that it is possible to find averaging methods that are less biased than FT, the mean weighted by discharges averaged over periods starting and ending between sampling times. In spite of these improvements, relatively large residual errors remained, even if these estimators were calculated over periods of 8 weeks. These methods should be tested further with solutes having different dependencies against the discharge.

One of the regressions gave unbiased flux estimates in one catchment (Erlenbach, R1), but this was never the case for both catchments at the same time. There is probably always some regression model that yields more or less unbiased predictions of fluxes; for Nitrex, this may be a sigmoidal curve. The critical point is finding the best model. In our example, it was not possible to choose according to the coefficient of determination. In addition, the best model for one catchment may be grossly wrong for another one. This is an evident lack of robustness. Even if a regression with a low bias can be found, the residual errors are always high, showing that the mechanisms controlling the water quality are more complex than a multiple regression model.

The relationship between discharge and electrical conductivity is relatively narrow and can essentially be described as a dilution of soil water by precipitation. Solutes measured in the runoff of the same catchments show weaker dependencies upon discharge, and plant nutrients generally show more seasonality. Nitrate concentrations, for example (Schleppi *et al.*, 2004), show flushing, preferential flow and dilution effects in the short term (with corresponding hysteresis), along with seasonality and effects of nitrogen deposition in the longer term. The correlation with the discharge even changes sign depending on the time scale considered. Dissolved organic matter, in contrast, is marked by flushing mechanisms and is positively correlated to the discharge (Hagedorn *et al.*, 2000). Depending on the shape of the discharge–concentration relationship, biases and residual errors of regressions will vary. In the catchments considered, the electrical conductivity and the geologically derived elements like calcium or magnesium are subjected to relatively simple control mechanisms. It can thus be considered a best case for applying regressions. Weaker regressions would be expected for catchments including groundwater bodies with substantial in- and ex-filtration. Furthermore, regressions cannot be expected to give better results for other solutes with more complex control mechanisms. Regression methods can, therefore, only be recommended with extreme caution. A graphical check with special attention to high-discharge data is, in any case, essential to avoid gross errors.

CONCLUSIONS

Because of highly variable discharges, sampling runoff water from small or very small catchments deserves particular attention. Grab samples essentially give a biased estimate of solute fluxes because they do not give enough weight to discharge peaks. Sampling discharge peaks alone or along with grab samples gives better estimates of fluxes, both in terms of systematic and residual errors, but no method can be considered as precise and *a priori* accurate. The same is true for flux estimates based on regressions or on different weighted averages. These methods are, therefore, rather a way to save a poor sampling scheme than to implement a good one. Only discharge-proportional sampling schemes allow for essentially unbiased flux estimates without requiring the chemical analysis of many samples.

ACKNOWLEDGEMENTS

This study was partly financed by the Swiss Federal Office for Education and Science (grant 94.0090) and by the Swiss National Science Foundation (grant 31-061959.00). We thank Melissa Swartz for improving the English of the manuscript.

REFERENCES

Bukaveckas PA, Likens GE, Winter TC, Buso DC. 1998. A comparison of methods for deriving solute flux rates using long-term data from streams in the Mirror Lake watershed. *Water, Air, and Soil Pollution* **105**: 277–293.

Burch H. 1994. Ein Rückblick auf die hydrologische Forschung der WSL im Alptal. *Beiträge zur Hydrologie der Schweiz* **35**: 18–33.

Church MR. 1997. Hydrochemistry of forested catchments. *Annual Review of Earth and Planetary Sciences* **25**: 23–59.

Coats R, Liu F, Goldman CR. 2002. A Monte Carlo test of load calculation methods, Lake Tahoe basin, California–Nevada. *Journal of the American Water Resources Association* **38**: 719–730.

Cohn TA. 1995. Recent advances in statistical methods for the estimation of sediment and nutrient transport in rivers. *Reviews of Geophysics* **33**: (Supplement): 1117–1123.

Cooper DM, Watts CD. 2002. A comparison of river load estimation techniques: application to dissolved organic carbon. *Environmetrics* **13**: 733–750.

Dann MS, Lynch JA, Corbett ES. 1986. Comparison of methods for estimating sulfate export from a forested watershed. *Journal of Environmental Quality* **15**: 140–145.

Davis JS. 1986. Improving information utilization of data from rivers and streams—the role of seasonal factors and annual periodicity in the variance of biogeochemical parameters. *Trends in Analytical Chemistry* **5**: 247–251.

Ferguson RI. 1986. River loads underestimated by rating curves. *Water Resources Research* **22**: 74–76.

Fredriksen RL. 1969. A battery powered proportional stream water sampler. *Water Resources Research* **5**: 1410–1413.

Hagedorn F, Kaiser K, Feyen H, Schleppi P. 2000. Effects of redox conditions and flow processes on the mobility of dissolved organic carbon and nitrogen in a forest soil. *Journal of Environmental Quality* **29**: 288–297.

Hagedorn F, Schleppi P, Bucher JB, Flühler H. 2001. Retention and leaching of elevated N deposition in a forested ecosystem with Gleysols. *Water, Air, and Soil Pollution* **129**: 119–142.

Johnson AH. 1979. Estimating solute transport in streams from grab samples. *Water Resources Research* **15**: 1224–1228.

Jordan TE, Correll DL, Weller DE. 1997. Relating nutrient discharges from watersheds to land use and streamflow variability. *Water Resources Research* **33**: 2591–2600.

Keller HM, Burch H, Guecheva M. 1989. The variability of water quality in a small mountainous region. In *Regional Characterization of Water Quality*, Ragone S (ed.). *IAHS* Publication No. 182. IAHS Press: Wallingford; 305–312.
Laxen DPH. 1977. A specific conductance method for quality control in water analysis. *Water Research* **11**: 91–94.
Likens GE, Bormann FH. 1995. *Biogeochemistry of a Forested Ecosystem*, 2nd edn. Springer: New York.
Line DE, Harman WA, Jennings GD. 1998. Comparing sampling schemes for monitoring pollutant export from a dairy pasture. *Journal of the American Water Resources Association* **34**: 1265–1273.
Moldan F, Wright RF. 1998. Episodic behaviour of nitrate in runoff during six years of nitrogen addition to the NITREX catchment at Gårdsjøn, Sweden. *Environmental Pollution* **102**(S1): 439–444.
Pacini N, Zobrist J, Ammann A, Gachter R. 1997. Water-quality surveillance of Swiss rivers. *Chimia* **51**: 929–934.
Preston SD, Bierman VJ, Silliman SE. 1992. Impact of flow variability on error in estimation of tributary mass loads. *Journal of Environmental Engineering, ASCE* **118**: 402–419.
Ranalli AJ. 1998. An evaluation of *in-situ* measurements of water temperature, specific conductance, and pH in low ionic strength streams. *Water, Air, and Soil Pollution* **104**: 423–441.
Robertson DM, Roerish ED. 1999. Influence of various water quality sampling strategies on load estimates for small streams. *Water Resources Research* **35**: 3747–3759.
Schleppi P, Muller N, Feyen H, Papritz A, Bucher JB, Flühler H. 1998. Nitrogen budgets of two small experimental forested catchments at Alptal, Switzerland. *Forest Ecology and Management* **101**: 177–185.
Schleppi P, Hagedorn F, Providoli I. 2004. Nitrate leaching from a mountain forest ecosystem with Gleysols subjected to experimentally increased N deposition. *Water, Air, and Soil Pollution Focus* **4**: 453–467.
Schleppi P, Waldner P, Stähli M. 2005. Errors of flux integration methods for solutes in grab samples of runoff water, as compared to flow-proportional sampling. *Journal of Hydrology* DOI:10.1016/j.jhydrol.2005.06.034.
Schwartz SS, Naiman DQ. 1999. Bias and variance of planning level estimates of pollutant loads. *Water Resources Research* **35**: 3475–3487.
St-Hilaire A, Caissie D, El-Jabi N, Morin G. 1998. Évaluation de l'applicabilité d'une méthode statistique aux variations saisonnières des relations concentration–débit sur un petit cours d'eau. *Revue des Sciences de l'Eau* **11**: 175–190.
Stone KC, Hunt PG, Novak JM, Johnson MH, Watts DW. 2000. Flow-proportional, time-composited, and grab sample estimation of nitrogen export from an eastern Coastal Plain watershed. *Transactions of the ASAE* **43**: 281–290.
Swistock BR, Edwards PJ, Wood F, DeWalle DR. 1997. Comparison of methods for calculating annual solute exports from six forested Appalachian watersheds. *Hydrological Processes* **11**: 655–669.
Thomas RB, Lewis J. 1993. A comparison of selection at list time and time-stratified sampling for estimating suspended sediment loads. *Water Resources Research* **29**: 1247–1256.
Van Sickle J, Wigington Jr PJ, Church MR. 1997. Estimation of episodic stream acidification based on monthly or annual sampling. *Journal of the American Water Resources Association* **33**: 359–366.
Walling DE, Webb BW. 1985. Estimating the discharge of contaminants to coastal waters by rivers: some cautionary comments. *Marine Pollution Bulletin* **16**: 488–492.

