



Errors of flux integration methods for solutes in grab samples of runoff water, as compared to flow-proportional sampling

Patrick Schleppe^{*}, Peter A. Waldner, Manfred Stähli

WSL-Swiss Federal Institute for Forest, Snow and Landscape Research, CH-8903 Birmensdorf, Switzerland

Received 11 October 2004; revised 14 June 2005; accepted 29 June 2005

Abstract

In a small mountain stream (Erlenbach, Alptal, central Switzerland), water samples were taken during 15 years by two different methods: weekly-bulked, discharge-proportional sampling and fortnightly grab (instantaneous) sampling. Compared with continuous, in-stream measurements of the electrical conductivity, the discharge-proportional sampling proved to give very accurate and precise estimates of solute fluxes. Relatively to discharge-proportional samples, integration methods based on the grab samples resulted in more or less biased fluxes of: Ca, Mg, K, Na, SO_4^{2-} , Cl^- , NO_3^- and electrical conductivity. Unweighted means were all positively biased due to concentrations decreasing with the discharge. Means weighted by discharges averaged fortnightly were less positively biased. On the other hand, weighing by the instantaneous discharge gave negative biases. Fluxes calculated from regressions of concentrations against discharge were also negatively biased, but the accuracy could partly be improved by including further factors into the regressions, especially the seasonality. If both grab and discharge-proportional samples are taken in parallel for some time, it is possible to calibrate the fluxes calculated from the first against the second. Calibration periods of 1–7 years were tested and found to improve the accuracy in 8 other years used for validation. Five years were sufficient to achieve unbiased regression estimates and a precision better than 10%. Replacing a less accurate grab sampling scheme by a discharge-proportional sampling should thus be done with a transition period during which both methods are run in parallel, 5 years being advisable for most solutes and for small streams.

© 2005 Elsevier Ltd All rights reserved.

Keywords: Water chemistry; Catchment hydrology; Runoff; Sampling; Bias; Accuracy; Precision; Solute fluxes

1. Introduction

Headwater catchments integrate biogeochemical processes over space: analysing stream water at the outlet gives information on processes in the whole

area. To calculate fluxes of suspended or dissolved substances, measured concentrations need to be integrated over time. This requires both an appropriate sampling scheme and a correct numerical integration of discharge and concentrations. Several studies have been published comparing different sampling schemes and/or integration methods, either in a mathematical approach (Cohn, 1995; Schwartz and Naiman, 1999) or based on empirical measurements (Johnson, 1979; Dolan et al., 1981; Walling and

^{*} Corresponding author. Tel.: +41 1 739 24 22, fax: +41 1 739 24 88.

E-mail address: schleppe@wsl.ch (P. Schleppe).

Webb, 1985; Dann et al., 1986; Preston et al., 1992; Thomas and Lewis, 1993; Swistock et al., 1997; Webb et al., 1997; Bukaveckas et al., 1998; Line et al., 1998; Robertson and Roerish, 1999; Smart et al., 1999; Stone et al., 2000; Horowitz et al., 2001; 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 standard deviation) of the estimates. Among the cited publications, however, there is no agreement at all on which sampling and integration method would consistently give the best flux estimates. The question is even more open for small streams than for large rivers because, as shown by Preston et al. (1992) and by Cohn (1995), flux estimates are often strongly biased when the discharge shows a high temporal variability.

In a previous paper, we used the electrical conductivity of stream water (measured with a 10-minute time resolution) to simulate different sampling strategies along with a few methods of flux integration (Schleppei et al., *in press*). These simulations were based on two years of measurements in a small and in a very small catchment (64 ha and 0.16 ha) in the Alptal valley, central Switzerland. In both cases, the electrical conductivity was dominated by the effect of calcium carbonate dissolved from the sub-soil. Only flow-proportional sampling schemes appeared essentially unbiased in these catchments, where the discharge varies over four orders of magnitude. All other methods were biased or not robust across catchments. Sampling at fixed intervals (grab sampling) was biased towards higher conductivities because of a negative correlation with discharge. Different peak sampling strategies produced variable (positive or negative) biases, but no method was unbiased simultaneously in both catchments. The same was true for integration methods based on grab sampling, including different regression models.

In the larger of these catchments (Erlenbach), grab samples and flow-proportional samples were analysed in parallel during 15 years. The measured solutes were: Ca, Mg, K, Na, SO_4^{2-} , Cl^- and NO_3^- , along with the electrical conductivity. The first goal of the present study is, based on these long-term data, to compare the accuracy and precision of different flux estimates and to validate our previous simulations. A second goal is to test if fluxes calculated from old grab samples can be corrected to become comparable to

a newly introduced flow-proportional sampling. From a practical point of view, the question is here, how many years both methods have been run in parallel in order to obtain corrected fluxes with an adequate precision.

2. Material and methods

2.1. Catchment and instrumentation

The Erlenbach catchment (64 ha) is located in the Alptal valley (central Switzerland), at 1100–1600 m above sea level (Burch, 1994). 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 shists. The soils are classified as umbric Gleysols. Because of its high clay content, the mineral soil has a very low permeability. The Erlenbach catchment is facing to the west, with an average slope of 20%. Up to 40% of the catchment is covered by forests and 60% by wet grasslands. The climate is cool and wet: 6 °C average temperature and 2200 mm precipitation per year (altitude of the weather station: 1200 m). The catchment is equipped with a V-notch weir to measure the water discharge on a 10-min time step (Burch, 1994).

2.2. Water sampling and analyses

During 15 years, from 1983 to 1997, individual water samples were taken automatically (LiquiBox sampler, Endress + Hauser, Reinach, Switzerland) for each 0.1 mm runoff depth of the Erlenbach. These samples were pooled to a flow-proportional sample, which was kept refrigerated (approx. 7 °C) and analysed weekly. In addition, grab samples were taken and analysed fortnightly. Analyses were done by inductively-coupled-plasma mass-spectrometry (ICP-MS) for the cations (Ca^{2+} , Mg^{2+} , K^+ , Na^+) and by ionic chromatography for the anions (NO_3^- , Cl^- , SO_4^{2-}). NO_3^- is the only solute considered here for which the atmospheric deposition is larger than the outflow, showing thus a net retention in the catchment (Schleppei et al., 1998). The electrical conductivity (adjusted to 20 °C) was measured in the water samples and also every 10 min directly in the stream. It was

shown to behave conservatively (linearly) upon mixing (Schleppe et al., in press) and can thus be integrated to fluxes in the same way as concentrations.

2.3. Methods of flux integration

Because chemical analyses are typically done at a much lower frequency than discharge measurements, concentrations have to be estimated either as an average over a period of sampling (average estimators) or for every time of discharge measurement (interpolation and regression methods).

2.3.1. Average estimators

Different methods were compared to estimate an average concentration directly from measurements in grab samples. The unweighted mean is calculated independently of the discharges (Eq. (1)) and corresponds to the method numbered 4 by Walling and Webb (1985). Means can also be weighted by the discharge: the instantaneous discharge at sampling time can be used (Eq. (2)), or an average discharge calculated for a period of time starting and ending in the middle between samples (Eq. (3)). These weighted means correspond to methods 5 and 3 of Walling and Webb (1985), respectively. They were calculated for each of the 15 years of sampling.

$$C = \frac{\sum C_i}{n} \quad (1)$$

$$C = \frac{\sum C_i \cdot Q_i}{\sum Q_i} \quad (2)$$

$$C = \frac{\sum C_i \cdot \bar{Q}_i}{\sum \bar{Q}_i} \quad (3)$$

where C = estimated average concentration, C_i = measured concentration, n = number of analyses, Q_i = instantaneous discharge at sampling time, and \bar{Q}_i = average discharge around sampling time (1 week before and 1 week after in the case of fortnightly samples).

In Eq. (2), concentrations can be considered as the quotient of fluxes divided by discharges. As a quotient, their estimate can be corrected according to Beale (1962). This Beale's ratio estimator (BRE, Eq. (4)) was also calculated for each of the 15 years.

As recommended by Preston et al. (1992), the data were first stratified according to the discharge at sampling time, building 3 strata per year.

$$C = \frac{\sum C_i Q_i \cdot \sum Q_i + n \cdot s_{CQ,Q}}{(\sum Q_i)^2 + n \cdot s_Q^2} \quad (4)$$

where $S_{CQ,Q}$ = covariance of fluxes ($C \cdot Q$) and discharges, and S_Q^2 = variance of discharges.

2.3.2. Interpolation

Another approach is to linearly interpolate concentrations between measurements (Eq. (5)). The interpolated values are then multiplied by the corresponding discharge measurements, and the obtained fluxes are integrated over time.

$$C(t) = \frac{C(t_i) \cdot (t_{i+1} - t) + C(t_{i+1}) \cdot (t - t_i)}{t_{i+1} - t_i} \quad (5)$$

where $C(t)$ = interpolated concentration, $C(t_i)$ and $C(t_{i+1})$ = measured concentrations and t = time, with $t_i < t < t_{i+1}$.

2.3.3. Regression methods

Regressions are commonly used to describe how concentrations vary with the discharge. Concentrations predicted by a regression model (in this case also called rating curve) are then used to estimate the solute fluxes. In 15 years, we collected a total of 407 grab samples, which could be used in the regressions. As usual for a small stream, the discharge of the Erlenbach has an approximately log-normal distribution (Keller et al., 1989) and we used thus its logarithm in all regressions. Three simple models were tested: a log-linear, a log-quadratic and a double-logarithmic regression. The log-linear and the log-quadratic models partly gave residuals which were not normally distributed. Further calculations were thus restricted to the double-logarithmic model (Eq. (6)), the residuals of which all had a positive kurtosis but were not significantly skewed and closer to normality than for the other models.

$$\log C = a + b \cdot \log Q \quad (6)$$

where C = concentration, Q = water discharge and a , b = regression parameters.

In this model, single concentrations are calculated by the exponentiation of the predicted log C , which

leads to a negative ‘retransformation bias’. We tested both the uncorrected regression and the correction proposed by Finney (1941) and adapted by Ferguson (1986) to logarithms in base 10 (Eq. (7)).

$$C' = C \cdot e^{k \cdot s^2} \tag{7}$$

where C' = corrected concentration, $k = (\ln 10)^2/2 = 2.65095$, and s = standard error of the estimated $\log C$.

Like Swistock et al. (1997), we also tried to include seasonality effects in the regression, using harmonic terms and their interaction with the discharge (Eq. (8)).

$$\log C = a + b \cdot \log Q + f(t) \tag{8}$$

where $t = 2\pi/365 \cdot \text{day of the year}$,

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

with c_1 to c_{12} = regression parameters.

Going backwards from the third harmonic, seasonality terms were pairwise removed from the models when neither the sine nor the cosine were significant.

Beside the discharge and the seasonality, other processes can affect concentrations in runoff water. Such changes can also be reflected in the electrical conductivity, and this parameter may thus be a useful information to predict the water chemistry. The seasonality term itself is an average over years and does not take the actual temperature into account. Since electrical conductivity and water temperature are often measured in the runoff, regression models including one or both of these variables were also tested (Eqs. (9–11)).

$$\log C = a + b \cdot \log Q + f(t) + d \cdot \log L \tag{9}$$

$$\log C = a + b \cdot \log Q + f(t) + e \cdot T \tag{10}$$

$$\log C = a + b \cdot \log Q + f(t) + d \cdot \log L + e \cdot T \tag{11}$$

where L = electrical conductivity, T = temperature and a to e = regression parameters.

2.3.4. Interpolation of residuals

With the regression methods, residuals (difference between observed and predicted values) are treated as errors: only the predicted values are used for the integration of fluxes over time. Another approach is to consider the residuals as a lack-of-fit, i.e. as an inability of the model to match the measured concentrations at sampling times. From this standpoint, it is possible to interpolate the residuals between the measurements (like in Eq. (5) for the concentrations) and to add them to the values predicted by regression. This method (Huntington et al., 1994) intends to take into account such changes which are affecting the concentrations between sampling times, but which cannot be adequately included in the regression. We tested the interpolation of residuals with the regression on discharge and seasonality (Eq. (8)).

2.4. Evaluation of the flux estimates

For the electrical conductivity, we integrated the measurements available at the same frequency as the discharge (10 min) and used this estimate as a reference. Flow-proportional sampling and estimates based on grab samples were compared to this reference. Differences were interpreted as consisting of a bias (systematic error) and a (random) standard deviation. The standard deviation was always calculated between years in order to enable comparisons also with average estimators, which were themselves calculated on a yearly basis.

For the analysed cations and anions, for which no continuous measurements were available, we chose as a reference the estimate, which would give the best results with the conductivity: the flow-proportional sampling.

2.5. Test of a change in the sampling method

In a long-term monitoring, grab samples may at some time be replaced by flow-proportional samples. This can introduce a shift in the time series and make it difficult to compare old

measurements with newer ones. If both sampling methods were used in parallel for some time, the direct comparison can be used to correct the old data and thus distinguish between true changes and methodological differences.

To test this procedure, we split our data into years used for calibration (i.e. to calculate a correction) and years used for validation (i.e. to check the accuracy of this correction). Flux estimates from grab samples were first compared to the fluxes from the proportional samples over the calibration dataset. A simple correction factor was calculated and applied to the validation dataset, indifferently on fluxes or on concentrations (Eq. (12)). Corrected estimates could then be compared with the actual fluxes from proportional samples taken during the validation period.

$$C' = C \cdot \frac{\bar{C}}{\hat{C}} \quad (12)$$

where C' = corrected concentration, C = concentration calculated from grab samples, \hat{C} = average concentration estimated from grab samples during calibration, \bar{C} = average concentration from proportional samples during calibration.

Using the same regressions for the whole dataset would eliminate possible changes of the discharge-to-concentration relationship with time, which can be important in some cases (Huntington et al., 1994). Here, regressions were thus calculated separately for the calibration and for the validation data. The goal of the procedure being a hindcast, the calibration period has to follow the validation. The calibration period was varied from 1 to 7 years to assess the effect of this duration on the obtained precision. The validation was always based on 8 years and thus longer than the calibration. Because no trend could be detected over the time in our measurements, we used all 15 years in rotation to test 15 starting points of the calibration. For example, a 4-year calibration starting in 1996 covered the years 1996, 1997, 1983 and 1984. The 15 starting points were treated as 15 replications of the procedure, even if they are not independent of each other.

3. Results

3.1. Discharge and concentrations

The grab samples taken during 15 years correspond to discharges covering 4 orders of magnitude, from 1.6 to 1600 l s⁻¹ km⁻². This range does not include the extreme values, since 0.09% of the measured discharges were between 1600 and the maximum runoff at 18,000 l s⁻¹ km⁻² (25 July 1984). However, the standard deviation of log Q is comparable between the sampling times and the entire dataset (0.66 vs. 0.62). The regressions were thus calculated on a broad data basis.

The measured concentrations and electrical conductivity all decrease with increasing runoff (Fig. 1). The solutes can be grouped as follows:

- Ca and Mg concentrations and the conductivity correlate well with the discharge, resulting in coefficients of determination $r^2 > 0.80$.
- Na and SO₄²⁻ concentrations plotted against log Q show a concave shape which is correctly fitted by the log–log model ($r^2 > 0.73$).
- Cl⁻ has a broader scatter, especially at the lower discharges, which is reflected in a lower r^2 of 0.33.
- K and especially NO₃⁻ show much scatter at all discharges. For NO₃⁻, the regression is able to predict only 7% of the observed variance.

3.2. Seasonality, conductivity and temperature effects

For Ca, Mg and electrical conductivity, the seasonality had a statistically highly significant but weak effect (Table 1). Values were higher in the spring than in the autumn, but only by a factor smaller than 1.2. Including the seasonality into the regression model improved their r^2 only marginally. All other concentrations showed a minimum in the spring and a maximum between late summer and winter. In terms of explained variance, the seasonality had a smaller effect than the discharge in the case of Na and SO₄²⁻. Here again, the r^2 values were not much improved. For K and Cl⁻, the seasonality had almost as much effect as the discharge, while NO₃⁻ concentrations were even more determined by the season than by the

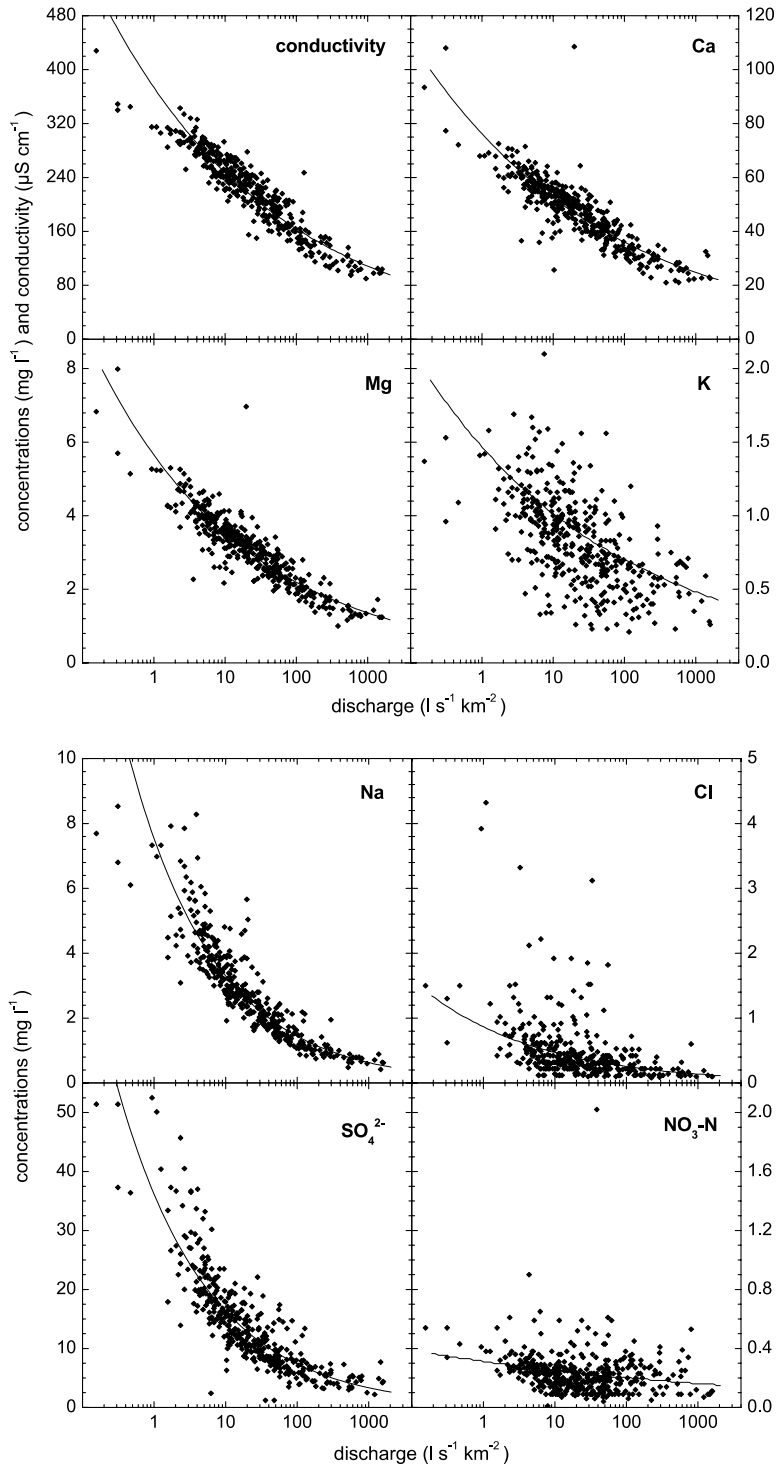


Fig. 1. Relationship between the logarithm of the discharge and concentrations of solutes in the Erlenbach stream. Data from 407 grab samples in 15 years. Curves are double-logarithmic regressions corrected for the retransformation bias.

Table 1

Parameters of the regressions of solute concentrations and electrical conductivity (L) in grab samples ($n=407$) against discharge (as logarithm), seasonality, water temperature and conductivity

Solute	r^2 of regressions based on...				Seasonality		
	Discharge	Discharge, seasonality	Discharge, seasonality, temperature	Discharge, seasonality, conductivity	Minimum	Maximum	Ratio max/min
L	0.88	0.92	0.92		December	May	1.18
Ca	0.80	0.84	0.84	0.86	September	March	1.09
Mg	0.85	0.87	0.88	0.89	September	May	1.13
K	0.34	0.46	0.47	0.46	April	October	1.54
Na	0.88	0.92	0.92	0.92	May	March	1.9
SO ₄ ²⁻	0.73	0.76	0.77	0.77	May	January	1.44
Cl ⁻	0.33	0.46	0.47	0.46	April	October	3.1
NO ₃ ⁻	0.07	0.27	0.27	0.29	May	August	2.7

All regressions are highly significant ($p < 0.0001$).

discharge. Considering the seasonality thus clearly improved these regressions in terms of r^2 .

The water temperature had a significant effect on all analysed parameters except Na and NO₃⁻. Compared to the regressions already including the seasonality, however, the additional variance explained by the temperature was always very small. The same is true for the conductivity: it was a significant predictor for all analysed concentrations but it barely improved the r^2 of the regressions.

3.3. Flow-proportional sampling

The electrical conductivity measured in flow-proportional samples was very close to the in-stream measurements integrated over the corresponding time periods (Fig. 2). The flux calculated from the samples was only 0.3% lower and thus practically unbiased. Very few points were below or above the ideal line and this scatter is reflected in a standard deviation of 0.3% on the yearly flux estimates.

3.4. Average estimates and interpolation

As expected from the negative correlations between discharge and concentrations, the calculation of unweighted means overestimated the fluxes of all solutes (Fig. 3). This bias was the strongest (close to 90%) for Na and SO₄²⁻, where low discharges tend to give much higher concentrations. Smaller but nevertheless significant biases (around 20%) were obtained

for those solutes with weaker relationships between discharge and concentration (K, Cl⁻, NO₃⁻). Standard deviations between yearly estimates, however, were larger for Cl⁻ and NO₃⁻ than for more discharge-dependent solutes.

Means weighted by the discharges at sampling time gave generally small negative biases and larger standard deviations. Taken individually, only the less discharge-dependent solutes were significantly biased

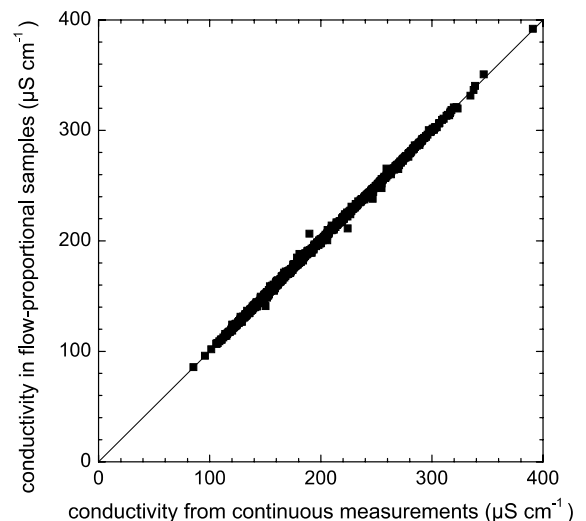


Fig. 2. Conductivity in flow-proportional samples compared to the integration of automatic measurements every 10 min ($n=758$, in 15 years).

(−18% for K and −29% for Cl[−]), but NO₃[−] was exceptionally accurate. Weighting the means by average discharges around sampling time again reversed the sign of the bias, becoming significantly positive except for Cl[−]. In all cases, this method was intermediate between the previous ones in terms of biases. Na and SO₄^{2−} showed again the strongest overestimations (~40%). According to the discharge-proportional samples, three solutes show important differences between years: K, Cl[−] and NO₃[−]. Some inter-annual variability is also observed with Na and SO₄^{2−}. Weighting by average discharge was in each case better able to reproduce these patterns than weighting by instantaneous discharge (Fig. 4). For these five solutes, there were significant correlations between measured and calculated values (Table 2). The Beale's ratio estimators had a positive bias intermediate between the unweighted mean and the mean weighted by average discharges. It was significant for all solutes except Cl[−] and NO₃[−]. The interpolation of concentrations between measurements, finally, gave results very similar to the mean weighted by average discharges, again with significant biases except for Cl[−] and NO₃[−].

3.5. Regression estimates

The regressions against the discharge gave negatively biased flux estimates (Fig. 5). The conductivity was not significantly biased and close correspondences were obtained also with Ca and Mg, while Cl[−] showed the strongest underestimation. The correction of Ferguson (Eq. (4)) reduced these underestimations. For the conductivity, Ca and NO₃[−], the biases were not significant; all other regression estimates were improved but remained negatively biased. Including the seasonality into the regression models had a small positive effect on the concentration and flux estimates. Thereby, the differences between solutes remained practically constant. The regressions with a small *r*² failed to reproduce the actual variability of the concentrations (K, Cl[−], NO₃[−]), giving flat patterns when estimated values are plotted against flow-proportional measurements (Fig. 6). Accordingly, these regression estimates were unable to reproduce the inter-annual variations: there were no significant correlations between predictions and discharge-proportional measurements on a yearly

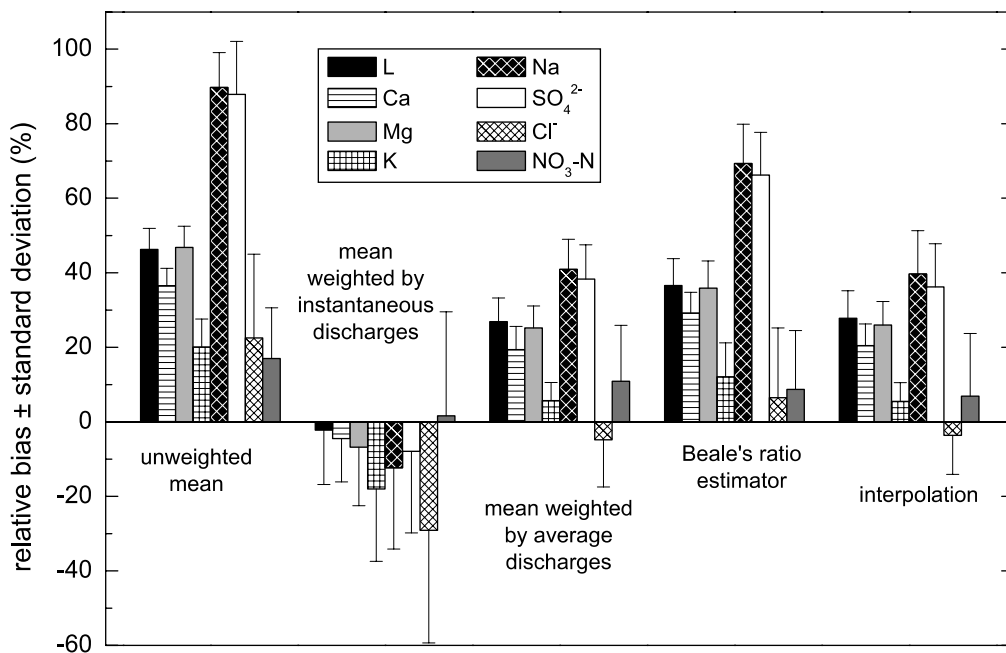


Fig. 3. Biases and standard deviations of annual flux estimates for different integration methods based on grab samples (*n* = 407, in 15 years), as compared to fluxes calculated from flow-proportional samples (*n* = 758).

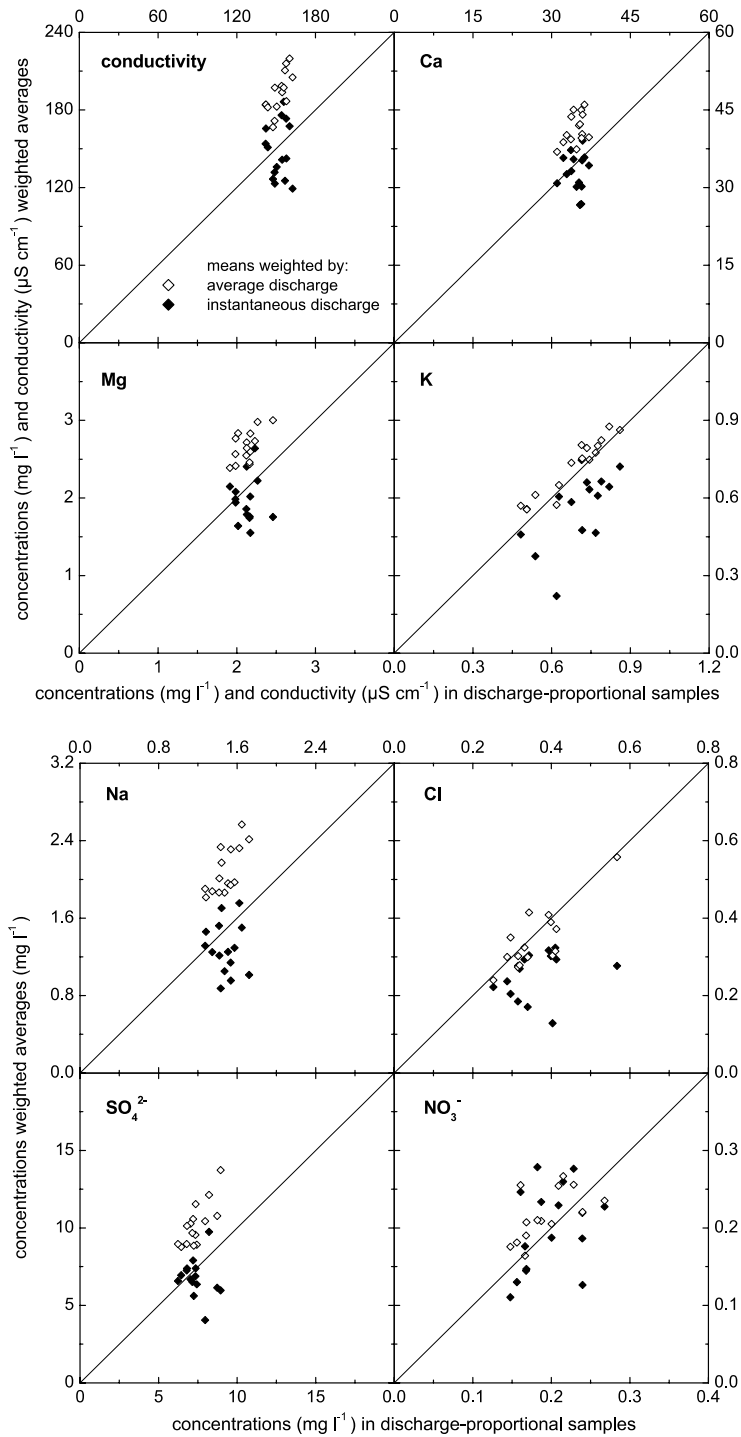


Fig. 4. Annual weighted means of solute concentrations in grab samples ($n=407$, in 15 years) as compared to concentrations calculated from discharge-proportional samples ($n=758$).

Table 2

Coefficients of correlation between yearly concentrations obtained by different integration methods and corresponding values obtained from discharge-proportional sampling ($n = 15$ years)

Integration method	L	Ca	Mg	K	Na	SO ₄ ²⁻	Cl ⁻	NO ₃ ⁻
Unweighted mean	0.65**	0.69**	0.67**	0.92***	0.68**	0.82***	0.78*	0.67**
Mean weighted by instantaneous discharges	0.02 ^{ns}	-0.01 ^{ns}	-0.03 ^{ns}	0.56 ^{ns}	-0.06 ^{ns}	-0.13 ^{ns}	0.28 ^{ns}	0.34 ^{ns}
Mean weighted by average discharges	0.71**	0.46 ^{ns}	0.60*	0.95***	0.71**	0.77**	0.82***	0.56*
Beale's ratio estimator	0.36 ^{ns}	0.48 ^{ns}	0.40 ^{ns}	0.88***	0.51*	0.75**	0.74**	0.65**
Interpolation	0.70**	0.54*	0.67**	0.95***	0.74**	0.76**	0.87***	0.54*
Regression (log Q)	0.59*	0.18 ^{ns}	0.29 ^{ns}	-0.10 ^{ns}	0.30 ^{ns}	0.34 ^{ns}	0.12 ^{ns}	0.09 ^{ns}
Regression ^a (log Q)	0.59*	0.18 ^{ns}	0.29 ^{ns}	-0.10 ^{ns}	0.30 ^{ns}	0.34 ^{ns}	0.12 ^{ns}	0.09 ^{ns}
Regression ^a (log Q , $f(t)$)	0.74**	0.42 ^{ns}	0.49 ^{ns}	-0.09 ^{ns}	0.42 ^{ns}	0.23 ^{ns}	-0.27 ^{ns}	0.31 ^{ns}
Regression ^a (log Q , $f(t)$, log L)	0.74**	0.42 ^{ns}	0.49 ^{ns}	-0.09 ^{ns}	0.42 ^{ns}	0.23 ^{ns}	-0.27 ^{ns}	0.31 ^{ns}
Regression ^a (log Q , $f(t)$, T)	0.76**	0.43 ^{ns}	0.50 ^{ns}	-0.06 ^{ns}	0.42 ^{ns}	0.21 ^{ns}	-0.27 ^{ns}	0.32 ^{ns}
Regression (log Q , $f(t)$) + interpolated residuals	0.84***	0.81 ^{ns}	0.93 ^{ns}	0.95 ^{ns}	0.84 ^{ns}	0.87***	0.86***	0.49 ^{ns}

L, conductivity; Q , discharge; $f(t)$, seasonality; T , water temperature. ns, not significant; *significant at $p=0.05$, **significant at $p=0.01$, ***significant at $p=0.001$.

^a Retransformation bias corrected (Ferguson, 1986).

basis (with the exception of the electrical conductivity, in spite of its narrow inter-annual variation range, Table 2). In general, the standard deviations of the regression estimates were roughly as large as the biases (in absolute values).

Considering also the water temperature and the electrical conductivity as regressors had only a very small effect on the estimated concentrations (Fig. 5), except for a slight improvement with NO₃ (effect of the temperature). Slightly less standard deviations for Ca

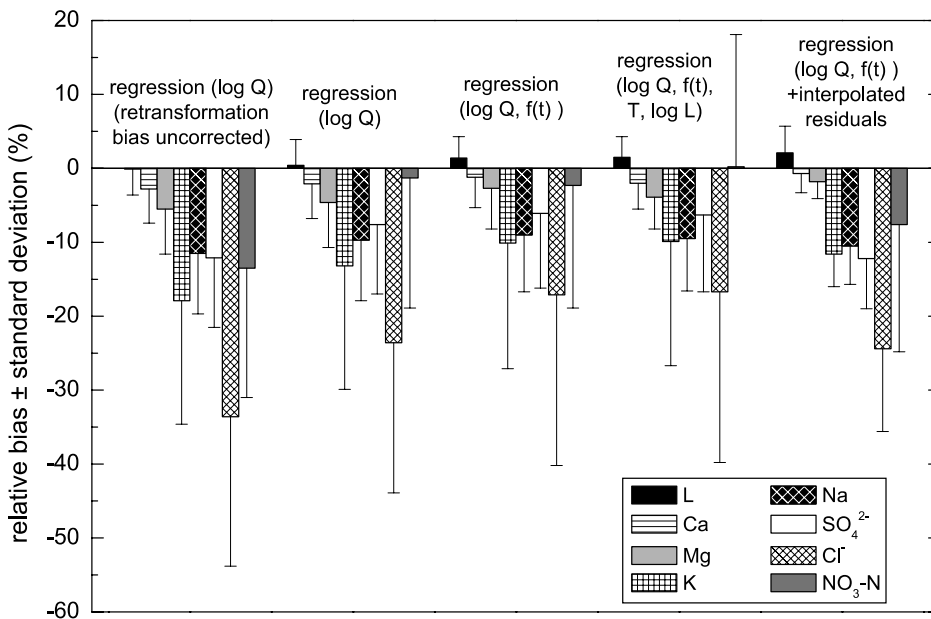


Fig. 5. Biases and standard deviations of annual flux estimates for different regression estimates based on grab samples ($n = 407$, in 15 years), as compared to fluxes calculated from discharge-proportional samples ($n = 758$). Q = discharge, $f(t)$ = seasonality, T = water temperature, L = electrical conductivity.

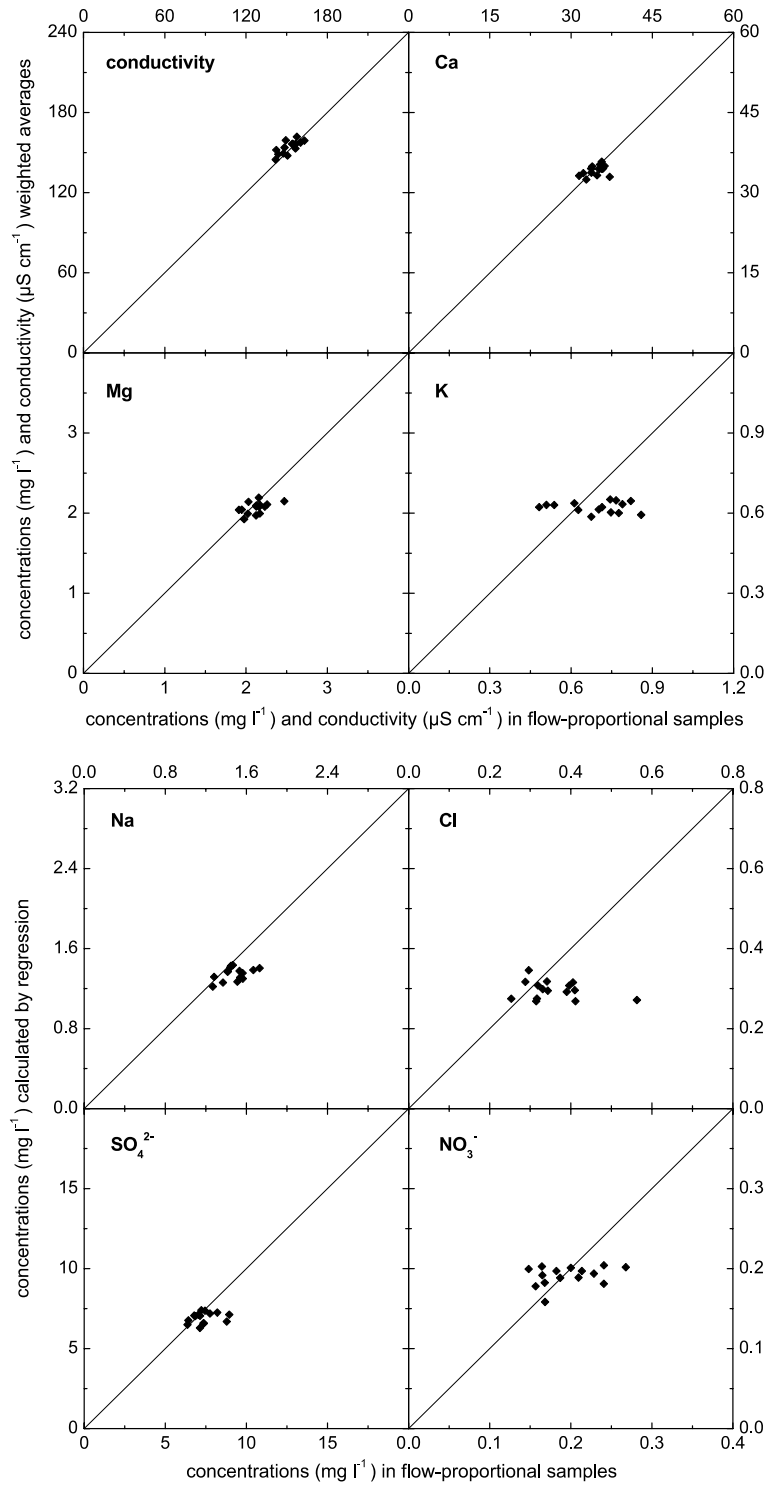


Fig. 6. Annual conductivity and concentrations estimated by regressions on discharge and seasonality, as compared to flow proportional samples ($n=758$, in 15 years).

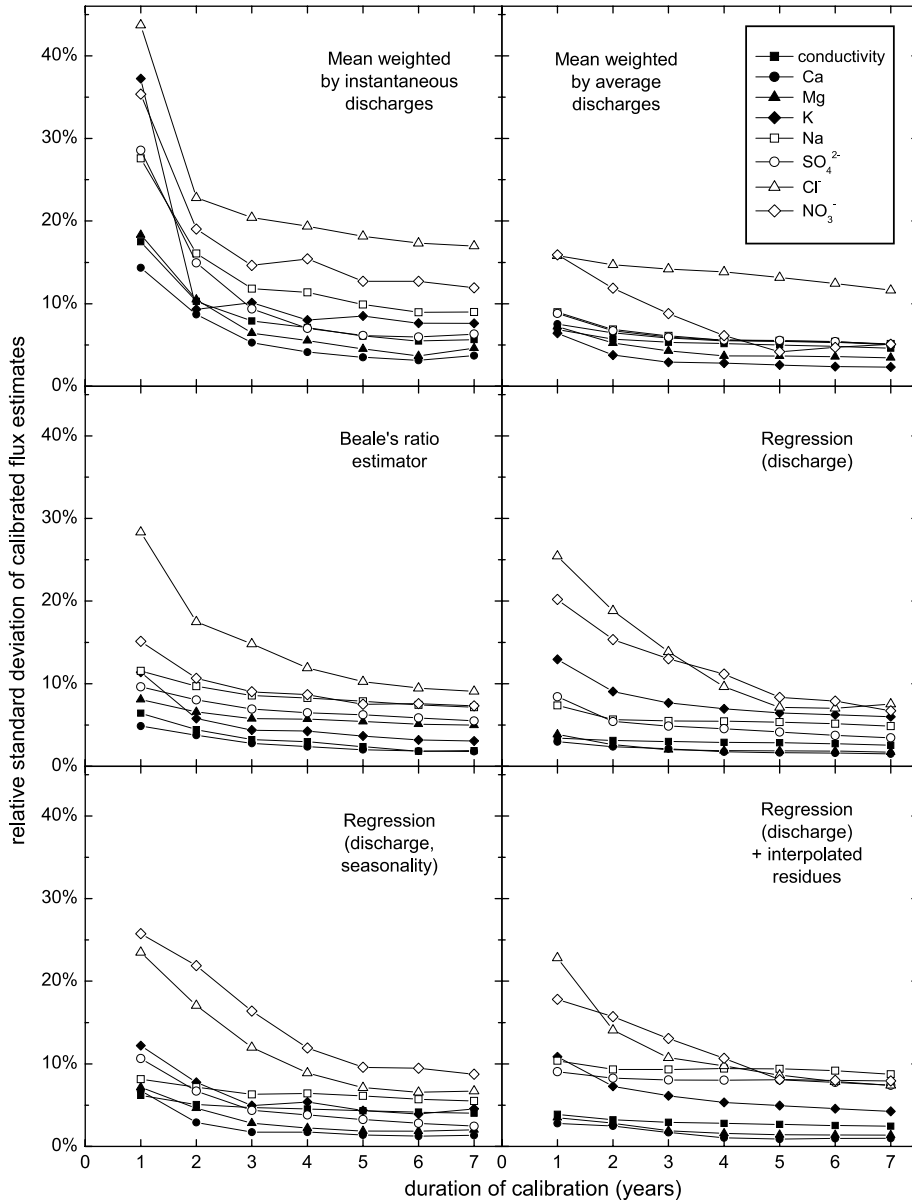


Fig. 7. Relative standard deviations of flux estimates with calibrations based on 1–7 years and applied to a 8-year validation periods. Integration methods based on grab samples, replicated for 15 starting years of calibration and compared with fluxes calculated from discharge-proportional samples.

and Mg were observed when the conductivity was included into the regression; accordingly, their yearly estimates became significantly correlated with the discharge-proportional sampling (Table 2).

Adding the interpolated residuals to the regression model with seasonality had only small effects on the

biases: there was a slight improvement for some solutes, but some worsening for others (Fig. 5). In most cases, standard deviations could be reduced. The inter-annual variations were much better represented by this model than by the regression itself: the correlations between annual concentrations obtained by this method and by

discharge-proportional sampling were high for all solutes except NO_3^- (Table 2).

3.6. Calibration of grab sampling against flow-proportional sampling

The biases shown in Figs. 3 and 5 were recalculated for calibration periods of 1–7 years. These numbers were then used to correct the estimates obtained from validation periods of 8 years. For every integration method and every solute, the corrected estimates were sometimes higher and sometimes lower than the flow-proportional estimates. In average of the 15 replications (=starting years of the calibration), the biases were within $\pm 10\%$ with a single year of calibration. With five years of calibration, all average biases were within $\pm 1\%$. Within the replications, however, the standard deviations (Fig. 7) were relatively large, but also decreased with the duration of calibration. Due to this variability, the biases were not significantly different from zero in any case, i.e. in any combination of solutes, integration methods and durations of calibration (T -tests, $0.15 < p \leq 1.00$). Cl^- and NO_3^- were more variable, while Ca, Mg and the conductivity gave more stable flux estimates. The Beale's ratio estimator and the different regression methods gave very similar standard deviations. They decreased with the number (n) of calibration years. The standard deviations were proportional to $1/\sqrt{n}$ with coefficients of determination $r^2 > 0.94$. The improved precision obtained with longer durations was thus as expected for an increasing sample size n .

The means weighted by instantaneous discharges produced larger standard deviations, roughly twice as large for the regressions. The results for a calibration of only one year were especially variable, with 5 solutes showing more than 25% standard deviation. An opposite behaviour was observed with means weighted by average discharges, with quite small standard deviations even with the shortest calibration. Both weighted means did therefore not follow the $1/\sqrt{n}$ rule.

4. Discussion

By adjusting its frequency, the discharge-proportional sampling gives the same weight to any

amount of runoff depth (to any 0.1 mm in our case). This way, the total fluxes are theoretically represented in the samples. By comparing the electrical conductivity in such samples with continuous in-stream measurements, we could verify that this method is essentially unbiased. The few points away from the ideal line in Fig. 2 could not be explained, but probably result from short unnoticed failures of the sampling device. Overall, however, both the method itself and its implementation in our catchment can be considered as giving highly accurate and precise results. This confirms results previously simulated for a shorter period of time (Schleppe et al., in press). Discharge-proportional sampling was therefore used as a reference to compare flux estimates based on grab (fixed-time) samples.

Calculating unweighted concentration averages gives the same importance to any sampling time. The periods of low discharge are then over-represented compared to their contribution to total solute fluxes. Because all solutes considered here are negatively correlated with the discharge, unweighted means are all overestimating the fluxes. As theoretically expected, this bias is the largest for those solutes with concentrations increasing the most at low discharges (Na, SO_4^{2-}); it is the lowest for solutes correlating only weakly with the discharge (K, Cl^- , NO_3^-). It is important to recall that this effect would be inverse for concentrations increasing with the discharge, as shown by numerous studies (e.g. Walling and Webb, 1985; Cooper and Watts, 2002) for suspended sediments, particulate substances and dissolved organic matter.

Means weighted by the discharge at sampling time gave negative biases, but it was the contrary for means weighted by the discharge in the period around sampling. This difference was already shown by Walling and Webb (1985). Taking the instantaneous discharge as a weighting factor allows to take short-term variations into account, i.e. concentrations changing with the discharge within a hydrological event. This method gives correspondingly good estimates (small negative biases) for those solutes which concentrations are closely related to the discharge. Because of this relative accuracy, this method is widely used. It must be noted, however, that the variability of the discharge is reflected as a standard deviation larger than with other estimates.

For this reason, the biases are often not significant. In other words, highly variable weighting factors are necessary for the accuracy, but have a negative effect on the precision. For K and Cl^- , with concentrations weakly related to the discharge, this gave significant biases. These solutes were better estimated by weighting with the average discharge around sampling time, two weeks in our case. This method apparently better takes slower concentration changes into account, i.e. changes effective over several hydrological events. Its intermediate biases between the unweighted mean and the mean weighted by instantaneous discharges are in line with the fact that the variance of the discharge is here taken into account between, but not within the sampling periods. The ability to reproduce yearly variations is an advantage of this method shown by our study.

Several authors found the stratified Beale's ratio estimator (BRE) to have small biases (Dolan et al., 1981; Preston et al., 1992; Young et al., 1988). These studies were based either on sediment or on phosphorus concentrations, both showing a positive correlation with discharge. In our case, all concentrations decreased with the discharge and BRE was stronger positively biased than the means weighted by average discharge. Nitrate was here an exception, but this may be due to the fact that the correlation between concentration and discharge changes sign depending on the considered time scale: positive within an event, negative over weeks or years (Schleppi et al., 2004). In all other cases, the correction brought by the BRE calculation was inappropriate and introduced a stronger bias as compared to the mean weighted by average discharge.

Graphically, the difference between the interpolation method and the mean weighted by average discharge lies in the integration of fluxes: as trapezes for the interpolation, as rectangles in the other case. This difference barely affected the accuracy and precision of the results. The interpolation has only the drawback of a somewhat more complicated computation, but with the advantage of giving a flux estimate for any period, independently of the sampling times.

In most cases, the regression methods underestimated the solute fluxes, but Ferguson's correction and the inclusion of the seasonality in the model improved the results. Further parameters

barely improved the biases. In an ideal logarithmic regression with Ferguson's correction, there is theoretically no bias. This means that the observed biases are due to a deviation from the ideal regression model. A reason may be the non-normality of the residuals or a lack-of-fit of the chosen function. The residuals of the log–log regressions did not show strong non-normalities or clear patterns, which should be interpreted as a lack-of-fit, but even an inconspicuous deviation from the model at the higher discharges may produce a bias. Another deviation from the theoretical regression model is the fact that the discharge and concentrations are time series, i.e. variables with auto- and cross-correlations. Discharge peaks usually produce hysteresis corresponding to loops in discharge-concentrations graphs. Regression methods applied to the idealised loops defined by Evans and Davies (1998), for example, can generate biases between approximately -4% and $+4\%$ (our calculations based on their published graphs). In the Erlenbach catchment, intensive sampling during and after a single rain event (Hagedorn et al., 2000) gives the following biases on fluxes estimated by regression: -47% (Ca), $+3\%$ (Cl^-) and -31% (NO_3^-). From these few examples, it is not possible to draw rules linking the hysteresis typology to the sign and magnitude of the bias, but is clear that strong and variable biases can be obtained. Because the hysteresis itself is a function of component mixing during runoff generation (Evans and Davies, 1998), this effect can vary between events, between solutes as well as between catchments.

The main disadvantage of regression models is that they completely eliminate any effect not explicitly considered, for example differences between years. Factors not accounted for may be the land-use, the weather conditions (including snow cover) not reflected in the discharge, the trends in climate or atmospheric deposition etc. Such effects are partly taken into account for NO_3^- when the water temperature is measured continuously and included into the regression. This temperature-dependency is likely a consequence of NO_3^- leaching being at least partly biologically controlled. Ca and Mg are the main cations and highly correlate with the conductivity ($r=0.90$ and 0.92 , respectively). This explains why the model including

the conductivity is better able to reproduce year-to-year variations. As previously shown by [Huntington et al. \(1994\)](#) for SO_4^{2-} (in a catchment with a net SO_4^{2-} retention), the addition of interpolated residuals is well able to follow inter-annual variations in solute fluxes. NO_3^- is here an exception. In our catchment, nitrate concentrations appear to be controlled by biological as well as by hydrological processes ([Schleppe et al., 2004](#)). Mechanisms like the flushing of NO_3^- produced by nitrification during a relatively dry period cannot be accounted by the regression model and are too short to show up in the interpolation of residuals. The dynamics of the retention and release processes is thus more important for the N budget of the catchment.

For all integration methods, an improvement of the accuracy can be achieved if estimates based on grab samples are calibrated against a discharge-proportional sampling. Within a long-term monitoring, it is thus possible to introduce a more accurate discharge-proportional sampling without losing the information gained in the previous years of grab sampling. As the obtained biases were not significantly different from zero, this approach can be considered as essentially accurate. Standard deviations were larger for solutes showing more year-to-year variations (Na, Cl^- , NO_3^-) but they decreased with the duration of the calibration. With the regression methods, dividing the errors by 2 requires to multiply the duration of the calibration by 4. Five years were sufficient to obtain standard deviations <10%. In long-term studies, an accuracy better than 10% is often difficult to maintain because of other errors arising from sample handling, storage and chemical analyses. With 5 years of calibration, the error due to the sampling scheme can thus be reduced to a level comparable with these other error sources. This duration can therefore be considered as an optimum for most solutes. Unlike the uncalibrated method, the calibrated regression was not more accurate when considering the seasonality or when interpolating the residuals. This means that the corrections brought by the calibration were more important than a more complex regression model. Beside regressions, means weighted by average discharges are also well responding to the calibration. On the contrary, because of their larger standard deviations, means weighted by instantaneous discharges cannot be recommended.

In our data, there was no long-term trend in any of the measured concentrations. We must therefore also ask if

a trend could change our rating of the calibration-validation approach. The means weighted by average discharges were already shown to be quite accurate, precise and able to reveal year-to-year changes, without requesting any explicit model for these changes. A trend would thus barely affect their accuracy. The same is true for the more complicated method of a regression plus interpolated residuals, with the residuals capturing the long-term trend. With the other regressions, models for the calibration and validation periods are calculated separately. A systematic difference between these periods would thus enter into the calculations and would not change the biases. Trends within each period, on the contrary, would not be taken into account and would increase the standard deviations. Based on these considerations, the choice of the integration model may be different from case to case.

5. Conclusions

Discharge-proportional water sampling is essentially unbiased. Weekly samples taken over 15 years showed us that this method gives highly accurate and precise measurements of solute fluxes even in a small, flashy mountain stream. Compared to discharge-proportional samples, all integrations based on grab samples are potentially biased and imprecise. The magnitude of the biases and standard deviations depends on the considered solute and on the chosen average or regression estimate. The relationship between discharge and concentration is a key factor affecting the accuracy and precision of these integration methods.

In small catchments where grab sampling was performed for several years, it is possible to introduce a discharge-proportional sampling to improve the accuracy of the flux estimates. Both sampling schemes should however be run in parallel for a few years because this will allow to partly correct for the biases associated with the old data. In most cases, 5 years of calibration appear to be a good compromise between expenses and obtained precision.

Acknowledgements

We express our gratitude to the late Hans M. Keller who established the study of the Erlenbach catchment

and lead it until 1992. We thank our past and present colleagues for their field work, sample preparation and data elaboration during the 15 years of this study: Thomas Hofstetter, Bruno Fritschi, Paul Weibel, Pierre Sulser, Karl Steiner, Hans Burch, Felix Forster and Peter Klöti. We also thank the team of the WSL central laboratory and their past and present leaders Maria Guecheva and Daniele Pezzota for the chemical analyses of our water samples. Statistical advice was given by Christian Hofmann and English corrections were kindly provided by Brian McArdell.

References

- Beale, E.M.L., 1962. Some uses of computers in operational research. *Industr. Org.* 31, 27–28.
- Bukaveckas, P.A., Likens, G.E., Winter, T.C., Buso, D.C., 1998. A comparison of methods for deriving solute flux rates using long-term data from streams in the Mirror Lake watershed. *Water Air Soil Pollut.* 105, 277–293.
- Burch, H., 1994. Ein Rückblick auf die hydrologische Forschung der WSL im Alptal. *Beitr. Hydrol. Schweiz* 35, 18–33.
- Coats, R., Liu, F., Goldman, C.R., 2002. A Monte Carlo test of load calculation methods, Lake Tahoe basin, California-Nevada. *J. Am. Water Resour. Assoc.* 38, 719–730.
- Cohn, T.A., 1995. Recent advances in statistical methods for the estimation of sediment and nutrient transport in rivers. *Rev. Geophys. Suppl.*, 1117–1123.
- Cooper, D.M., Watts, C.D., 2002. A comparison of river load estimation techniques: application to dissolved organic carbon. *Environmetrics* 13, 733–750.
- Dann, M.S., Lynch, J.A., Corbett, E.S., 1986. Comparison of methods for estimating sulfate export from a forested watershed. *J. Environ. Qual.* 15, 140–145.
- Dolan, D.M., Yui, A.K., Geist, R.D., 1981. Evaluation of river load estimation methods for total phosphorus. *J. Great Lakes Res.* 7, 207–214.
- Evans, C., Davies, T.D., 1998. Causes of concentration/discharge hysteresis and its potential as a tool for analysis of episode hydrochemistry. *Water Resour. Res.* 34, 129–137.
- Ferguson, R.I., 1986. River loads underestimated by rating curves. *Water Resour. Res.* 22, 74–76.
- Finney, D.J., 1941. On the distribution of a variable whose logarithm is normally distributed. *J. R. Stat. Soc.*, B 7, 155–161.
- Hagedorn, F., Schleppei, P., Waldner, P., Flühler, H., 2000. Export of dissolved organic carbon and nitrogen from Gleysol dominated catchments - the significance of water flow paths. *Biogeochemistry* 50, 137–161.
- Horowitz, A.J., Elrick, K.A., Smith, J.J., 2001. Estimating suspended sediment and trace element fluxes in large river basins: methodological considerations as applied to the NASQAN programme. *Hydrol. Process.* 15, 1107–1132.
- Huntington, T.G., Hooper, R.P., Aulenbach, B.T., 1994. Hydrologic processes controlling sulfate mobility in a small forested watershed. *Water Resour. Res.* 30, 283–295.
- Johnson, A.H., 1979. Estimating solute transport in streams from grab samples. *Water Resour. Res.* 15, 1224–1228.
- Keller, H.M., Burch, H., Guecheva, M., 1989. The variability of water quality in a small mountainous region. *IAHS Publ.* 182, 305–312.
- Line, D.E., Harman, W.A., Jennings, G.D., 1998. Comparing sampling schemes for monitoring pollutant export from a dairy pasture. *J. Am. Water Resour. Assoc.* 34, 1265–1273.
- Preston, S.D., Bierman, V.J., Silliman, S.E., 1992. Impact of flow variability on error in estimation of tributary mass loads. *J. Env. Eng. ASCE* 118, 402–419.
- Robertson, D.M., Roerish, E.D., 1999. Influence of various water quality sampling strategies on load estimates for small streams. *Water Resour. Res.* 35, 3747–3759.
- Schleppei, P., Muller, N., Feyen, H., Papritz, A., Bucher, J.B., Flühler, H., 1998. Nitrogen budgets of two small experimental forested catchments at Alptal, Switzerland. *For. Ecol. Manage.* 101, 177–185.
- Schleppei, P., Hagedorn, F., Providoli, I., 2004. Nitrate leaching from a mountain forest ecosystem with Gleysols subjected to experimentally increased N deposition. *Water Air Soil Pollut. Focus* 4, 453–467.
- Schleppei, P., Waldner, P.A., Fritschi, B. in press. Accuracy and precision of different sampling strategies and flux integration methods for runoff water. *Hydrol. Proc.*
- Schwartz, S.S., Naiman, D.Q., 1999. Bias and variance of planning level estimates of pollutant loads. *Water Resour. Res.* 35, 3475–3487.
- Smart, T.S., Hirst, D.J., Elston, D.A., 1999. Methods for estimating loads transported by rivers. *Hydrol. Earth Syst. Sci.* 3, 295–303.
- Stone, K.C., Hunt, P.G., Novak, J.M., Johnson, M.H., Watts, D.W., 2000. Flow-proportional, time-composited, and grab sample estimation of nitrogen export from an eastern Coastal Plain watershed. *Trans. ASAE* 43, 281–290.
- Swistock, B.R., Edwards, P.J., Wood, F., DeWalle, D.R., 1997. Comparison of methods for calculating annual solute exports from six forested Appalachian watersheds. *Hydrol. Process.* 11, 655–669.
- Thomas, R.B., Lewis, J., 1993. A comparison of selection at list time and time-stratified sampling for estimating suspended sediment loads. *Water Resour. Res.* 29, 1247–1256.
- Walling, D.E., Webb, B.W., 1985. Estimating the discharge of contaminants to coastal waters by rivers: Some cautionary comments. *Mar. Poll. Bull.* 16, 488–492.
- Webb, B.W., Phillips, J.M., Walling, D.E., Littlewood, I.G., Watts, C.D., Leeks, G.J.L., 1997. Load estimation methodologies for British rivers and their relevance to the LOIS RACS(R) programme. *Sci. Tot. Environ.* 194, 379–389.
- Young, T.C., Depinto, J.V., Heidtke, T.M., 1988. Factors affecting the efficiency of some estimators of fluvial total phosphorus load. *Water Resour. Res.* 24, 1535–1540.