Uncertainty in acquiring elemental fluxes from subtropical mountainous rivers

T. Y. Lee¹,², J. C. Huang³, A. E. Carey⁴, S. C. Hsu¹, K. Selvaraj¹, and S. J. Kao¹,⁵

¹Research Center for Environmental Changes, Academia Sinica, Nangang, Taipei, Taiwan
²Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei, Taiwan
³Department of Geography, Chinese Culture University, Taipei, Taiwan
⁴School of Earth Sciences and Byrd Polar Research Center, The Ohio State University, Columbus, Ohio, USA
⁵State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, China

Received: 25 October 2009 – Accepted: 16 November 2009 – Published: 2 December 2009
Correspondence to: S. J. Kao (sjkao@gate.sinica.edu.tw)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

Mountainous watersheds in high standing islands of the western tropical and subtropical Pacific have received great international attention regarding its high physical and chemical weathering rates caused by cyclone invasion, friable lithology and high tectonic activity. Since mountainous region is usually difficult to assess, particularly, during severe weather conditions, hydrological responses of elements against full-scale of water discharge (often >2 orders of magnitude) are rarely reported. In this study, we conducted discrete sampling (~3 day interval) throughout four seasons and intensive sampling (hourly) during typhoon floods from three upstream watersheds in Taiwan during 2002–2005. These observations revealing various elemental responses are taken as complete dataset (i.e. reference flux) to evaluate flux uncertainty among constituents caused by different sampling frequency, sample size and estimators. Five constituents are analyzed, including nitrate (NO₃), sulfate (SO₄), dissolved organic carbon (DOC), calcium (Ca), and silicate (Si). Each has specific environmental and geological implications. Direct average, flow-weighted and rating curve methods were applied. Basing on statistical analyses, flow-weighted method is the most conservative method, and is recommended to use for all constituents if few samples are available. The rating curve method is suggested, only when water samples in high-flows are available. Direct average method is only appropriate for stable constituents, such as Si. These findings, such as concentration-discharge variation, sampling frequency effect, and flux estimator assessment, offer fundamental knowledge while estimating geochemical fluxes from small mountainous rivers in Oceania region.

1 Introduction

Oceania Rivers in the tropics and subtropics are characterized by steep slope, small watershed area, episodic intense precipitation and high water runoff; consequently, holding high nutrient fluxes, physical and chemical erosion rates. Collectively, Oceania
Rivers account for 12% of the global water discharge (Milliman, 1991; Milliman et al., 1999) playing an important role in global geochemical fluxes (Kao and Liu, 1996, 1997; Lyons et al., 2002, 2005; Goldsmith et al., 2008). Estimating reliable geochemical fluxes for those small mountainous rivers (SMRs) are therefore crucial in estimation of global elemental budget. However, to monitor the outlet at downstream for entire watershed is not easy due to episodicity of water discharge (Kao and Milliman, 2008); not mention to carry out intensive sampling at pristine upstream that is difficult to reach due to traffic inconvenience particularly during cyclone invasion. Install instruments to sample water continuously is impossible due to extreme hydropower, highly fluctuated water level and rolling gravels (even bridge can be washed away) during typhoon periods. Despite of difficulties, to obtain the background flux (without land use and perturbation) from upstream is essential to understand unperturbed drainage system and to differentiate anthropogenic impacts from natural elemental cycle. Without such baseline information, input-output mass balance approach (Butturini, 2005) and cumulative flux approach (Neal, 2002) to determine solute detention/release in stream reaches or to distinguish differences between several watersheds with different anthropogenic disturbances can not be achieved. So far, there are no studies discussing about the uncertainties of flux estimates for Oceania geographical settings, moreover, uncertainties associated with flux estimates based on infrequent samples increase with decreasing the river size (Rode and Suhr, 2007). Above reasons underscore the importance to quantify uncertainties in acquiring elemental fluxes from subtropical mountainous rivers.

In previous studies, intensive sampling scheme, stratified sampling or fixed sampling interval schemes (Kronvang and Bruhn, 1996; Robertson and Roerish, 1999; Thomas and Lewis, 1993, 1995) were widely discussed. However, their sampling schemes were un-doable due to the mentioned reasons above. For Taiwan rivers and other SMRs in Oceania, the most common impedances are road damages caused by landslide, rock fall and under-scouring due to extreme rainfall and stream power. Random sampling procedure will be the case that most researchers encounter in such
environment. Therefore, properly transfer discrete yet valuable concentration data to flux within specific time span by using either continuous or average water discharge records is state-of-art (Smart et al., 1999). Moreover, elements may reveal distinctive hydrological responses (enhancement and dilution) to complicate flux estimation. Empirical (Walling and Webb, 1981; Littlewood, 1995), theoretical and statistical methods (Ferguson, 1987; Clarke, 1990) have been proposed to assess uncertainties associated with flux estimation due to under-sampling. Of which, suspended sediment loads (Walling and Webb, 1981; Crawford, 1991; Philipps et al., 1999; Horowitz, 2003; Coynel et al., 2004) had been thoroughly discussed; however, relatively fewer evaluations of nutrient load estimation were published because nutrient concentrations are less surveyed at higher sampling frequency (i.e. daily or higher).

For nitrate, Littlewood (1995) investigated the performance of two mass-load algorithms and sampling frequencies based on a generated time series of synthetic concentrations for two hypothetical behaviors, i.e. increasing and decreasing concentration with water discharge rate. Moatar and Meybeck (2005) compared performances of different algorithms for estimating annual nutrient loads in River Loire, France. The extent of errors in response to different calculation methods and different constituents, such as nitrate, total phosphate, orthophosphate, and particulate phosphate, were illustrated based on a monthly sampling scheme. Some studies proposed practical frameworks to choose the most accurate method basing on available data and analyzed the influence of sampling schemes on flux calculation (Quilbe et al., 2006; Kronvang and Bruhn, 1996; Robertson and Roerish, 1999; Thomas and Lewis, 1993, 1995). Nevertheless, most of them designed the sampling schemes in terms of fixed sampling interval, such as weekly, biweekly, monthly and sometimes event flows were supplemented.

In this study, we focus on five constituents, including nitrate (NO$_3$), sulfate (SO$_4$), dissolved organic carbon (DOC), calcium (Ca), and silicate (Si). Each of them has specific environmental and geological implications and thus can serve as an indicator or a tool to constrain nutrient emission (NO$_3$), non-point source pollutant (NO$_3$ and SO$_4$), chemical weathering and groundwater discharge (SO$_4$, Ca and Si), and organic
matter export (DOC). To identify the possible error distribution in flux estimation for SMRs due to under-sampling, we carried out observations to obtain representative range of elements concentration and their full-range hydrological responses aiming to investigate how estimation error responds to infrequent sampling scheme. Such entire spectrum of hydrological conditions would benefits for biogeochemical studies to completely describe the biogeochemical functioning of stream ecosystem (Butturini, 2005). Three commonly used methods, direct average, flow-weighted, and rating curve, were applied and evaluated. Results project error ranges of geochemical fluxes attributed to limited samples. Error range and distribution are useful reference for future studies in analogues regions and potentially applicable to previously reported values. Appreciation in various hydrological responses among elements may further help us to set proper sampling strategy on target element in subtropical mountainous rivers.

2 Study site

The Danshuei drainage basin (Fig. 1), located in northwestern Taiwan, is composed of three major tributaries, Keelung, Sindian, and Dahan Rivers. The basement rocks are mainly Tertiary argillite-slate and meta-sandstone (Ho, 1975). To determine uncertainties of geochemical flux estimates from pristine regions, three stations located at the upstream area of each tributary with population density less than 100 persons per km$^2$ were selected.

Three stations are Shioluan in the upstream of Dahan River (D), Diyu in the Sindian River (S), and Dongshe in the Keelung River (K), respectively (Fig. 1). The watershed areas of these three sampling stations are 78, 116 and 7 km$^2$, respectively. The elevation of stations are 827, 200, 220 m, and maximum elevations for the three watersheds are 2083, 829, and 540 m, respectively, above sea level. The annual rainfall ranges from 2000 to 4000 mm/yr over the entire watershed, with a mean of 3000 mm/yr (Hydrological Year Book of Taiwan). The dry season is usually from April to June. During summer, from late June to September, tropical storms (typhoons) cause torrential rain,
which typically accounts for 50% of the annual rainfall and episodic rainfall can be as high as 1000 mm/day. All three watersheds are relatively pristine and dominated by evergreen forests.

The corresponding daily water discharge rate on sampling days and hourly discharge rate during typhoons were taken from the Water Resources Agency (WRA) of Taiwan. The reported daily water discharge rate range from 2002 to 2005 respectively are 0.01–363, 0.86–187, and 0.001–8.55 m$^3$/s. The coefficients of variation of daily water discharge rate during these three years are quite similar (S:2.7; K:2.1; and D:2.5). However, the frequency distributions of daily discharge among the three stations are different reflecting the natural hydrological characteristics of respective sub-watershed (Fig. 2).

3 Material and methods

3.1 Sampling and measurements

Upon field sampling, discrete water samples were collected at 3-day interval from April 2002 through April 2005 (37 months) at stations D and K and up to July 2005 (40 months) at station S (Table 1). More intensive sampling (i.e. interval between two sampling is about 3 hours) at station K, the most pristine among the three watersheds, was conducted during four typhoons invading Taiwan in 2002 and 2003.

Upon collection, water samples were filtered through GF/F filters (0.7 µm) and the filtrate is separated into two aliquots. One aliquot was preserved at room temperature for Si, Ca and SO$_4$-S analyses and another was quick-frozen in liquid nitrogen for DOC and NO$_3$-N measurements.

Nitrate was determined by ion chromatography (IC) using either a Dionex ICS-90 or 1500 instrument using the method of Welch et al. (1996) with detection limit of 0.01 mg/l. DOC was measured by wet chemical oxidation with an OI-Analytical 1110 TOC analyzer (precision of ±5%). Calcium, silicate and sulfate were determined by
inductively coupled plasma-optical emission spectrometry (ICP-OES) with detection limit of 0.01 mg/l.

Here we give one example for Station K at Keelung River in Fig. 3 to illustrate the temporal dynamics of water discharge, SO$_4$-S, NO$_3$-N and DOC. The water discharge, apparently, is much variable compared to the concentration variability of SO$_4$-S, NO$_3$-N and DOC. There is a clear dilution phenomenon for SO$_4$-S during peak water discharge (e.g. indicated by red arrows) and its concentration increases as the water discharge decreases after the peak. However, patterns for NO$_3$-N (positively-correlated to water discharge) and DOC (discharge-independent) are totally different from that of SO$_4$-S (Fig. 4). Hence, it is interesting and important to realize how sampling strategies might affect flux estimation for every different constituent.

3.2 Estimator overview

Flux estimation is important due to the urgent demand from environmental management and geochemical studies. Over 10 estimators had been proposed during the past 2 decades. Those various estimators could be classified into direct average, flow-weighted and rating curve methods (Preston et al., 1988; Kronvang and Bruhn, 1996; Rode and Suhr, 2007). Many previous studies designed sophisticated method to assess the uncertainties under pre-designed sampling protocols. By contrast, we attempted to demonstrate the essential inter-correlation between sampling number (sample size) and estimator under the imperfect sampling condition for the mountainous area in Oceania region. The three methods are described below.

3.2.1 Direct average method

The direct average (DA) method is expressed as:

\[
\text{Load} = K \frac{1}{n} \sum_{i=1}^{n} C_i \times Q_t
\]  

(1)
where $K$ is the conversion factor to convert the calculated values into loads of specific period. $C_i$ (in mg/l) is the constituent concentration of $i$-th sample. $Q_i$ (in m$^3$/s) is total stream flow during the period concerned and $n$ represents the total number of samples. The hydrological response apparently is not considered in this estimator. Another motivation of using DA method is that remote area may not have flow gauge station. Therefore, long-term mean water discharge rate might be estimated by long-term mean precipitation and/or interpolated from downstream gauges by using area ratio.

### 3.2.2 Flow-weighted method

The flow-weighted (FW) method considers the hydrological response. When the corresponding (instantaneous) water discharge rate at the sampling time is available, the constituent concentration can be weighted according to instantaneous water discharge rate using the formula:

$$\text{Load} = K \frac{\sum_{i=1}^{n} C_i Q_i}{\sum_{i=1}^{n} Q_i} \times Q_t$$

(2)

where $Q_i$ (m$^3$/s) is the instantaneous water discharge rate at the sampling time. There are some modifications for FW method, such as ratio estimator, however, they are generally thought of as one of the principal alternatives to the rating curve method (Coats et al., 2002; Dolan et al., 1981). Nevertheless, the FW method gives consistent performance as ratio estimators (Cooper and Watts, 2002; Preston et al., 1988); hence, we only discuss the general form of FW method in the paper as Eq. (2) shows.
3.2.3 Rating curve method

The rating curve (RC) method is initially applied by constructing a power function relationship between the sampled constituent concentration and the instantaneous water discharge rate. Then, constituent concentrations on non-sampling days are assumed to follow the same power function, which is expressed as:

\[
\text{Load} = K \sum_{j=1}^{T} Q_j C_j = K \sum_{j=1}^{T} a Q_j^{b+1}
\]

(3)

where \(Q_j (m^3/s)\) is the daily water discharge rate, \(C_j (mg/l)\) is an estimated constituent concentration on the \(j\)-th day, \(T\) (days) is number of days of the period of concern, and, \(a\) and \(b\) are coefficients of power function and derived from observed concentration and water discharge rate by log-linear least-square method. Previous study by Kao et al. (2005) in sediment flux estimation for Taiwan Rivers showed that the back transformation from log-space to arithmetic-space may result in either underestimation or overestimation. Therefore, the correction factor proposed by Kao et al. (2005) is applied when solute rating curves are fitted by least-squares regression of logarithmically transformed data.

3.3 Statistical analysis

3.3.1 Reference flux

The reference flux provides the standard to examine the goodness of estimation. Previous studies usually set annual flux as the reference flux (e.g. Quilbe et al., 2006). However, as mentioned earlier, in many cases it is impossible to measure concentrations continuously, particularly, for those SMRs. Some studies (e.g. Moatar and Meybeck, 2005) linearly interpolate the concentrations measured on two adjacent sampling days to represent the constituent concentration for non-sampling days. While some other
studies gave values for non-sampling days by using flow-weighted method or log-linear regression method. In this study, we abandoned interpolation to avoid introducing any artificial uncertainty into the reference flux and prohibit benefit from any estimation methods prior to our data analysis. Hence, we calculated the reference flux by sum- 
marizing all observed daily fluxes.

To demonstrate the representativeness of reference flux, we compared frequency distribution of water discharge between our sampling days and the long-term record (Fig. 2). The frequency distribution of our 3-year sampling resembles that of long-term record (except station D during high flow), indicating a proper sampling of full spectrum of natural variability in our observation. For this reason, the observed cumulative load during the study period is reasonably taken as the reference flux to examine the performance of various sampling strategies.

3.3.2 Monte Carlo simulation

This method is applied to examine the effects of both the sample size/frequency and the selected estimator on the uncertainty of flux estimations (Guo et al., 2002). All fluxes estimated by three methods basing on sub-group samples were compared with the reference flux for performance evaluation. To simulate wide range sampling possibilities, sub-group samples are repeatedly extracted 10 000 times by random selection at fixed sample size from the original sample pool without replacement. Estimated fluxes derived from the 10 000 sets are analyzed. The sample size is from 10 to the maxima number (depending on observations) with fixed increment of 10. Through this random re-sampling procedure, the uncertainty caused by different sampling size and estimator can be obtained. Statistically, 10 000 sets of sub-group samples should reach the maxima range for error distribution, which potentially includes any sampling combination, even time-stratified sampling.
3.3.3 Root mean squared error (RMSE)

Each flux estimated from the sub-group samples is compared with the reference flux to determine the relative error (calculated as percentage) of the reference flux. The distribution pattern of these relative errors, ε, provides indices of accuracy and precision. The mean of the relative errors, $\bar{\varepsilon}$, is taken as an index for accuracy, and the standard deviation, $s$, serves as an index of precision. Previous researchers have proposed the RMSE as an evaluation criterion, which combines bias and precision, and calculated as follows (Walling and Webb, 1981; Dolan et al., 1981; Moatar and Meybeck, 2005):

$$\text{RMSE} = \sqrt{\bar{\varepsilon}^2 + s^2}$$

We also measure the efficiency of sampling by using the non-parametric Rank-Sum Test. The purpose is to determine the threshold while further increasing sample size may be ineffective. This threshold benefits sampling strategy in terms of time, cost and manpower savings.

4 Results

Figure 4 illustrates the range of each elemental concentration versus corresponding daily water discharge rate ($C-Q$ relationship) during 3-year sampling period. For samples collected during typhoon flood, instantaneous water discharge rate is applied. As expected, different elements show distinctive hydrological responses due to different biotic and abiotic processes occurred within respective watersheds.

The Ca concentrations at D (range between 7.7 and 38.4 mg/l with most values higher than 20 mg/l) are higher than those observed at station S and K, which fall in the ranges of 2.6–21.1 and 3.5–13.8 mg/l, respectively. Except the station D, Ca concentrations decrease as increasing water flow. Runoff dilution obviously plays a role. Coefficients of variation of Ca at stations S, K, and D are 0.43, 0.24, and 0.18, respectively. Larger variability in concentration tends to occur at lower flow.
The ranges of DOC concentration at stations S, K, and D are 0.35–2.86, 0.40–2.18, and 0.34–9.42 mg/l, respectively. Those ranges resemble that reported in Lanyang-Hsi River at north eastern Taiwan (0.5–4 mg/l). For DOC, coefficients of variation at stations S, K, and D are 0.39, 0.4, and 0.73, respectively. Water discharge seems not a key factor in controlling DOC concentration. Only slightly inverse relationships can be found at station S. (Note we failed to measure DOC collected during typhoon period because of instrumental problem).

The ranges of NO$_3$-N concentration for stations S, K, and D are 0.17–1.39, 0.11–0.77, and 0.03–0.81 mg/l, respectively. The observed NO$_3$-N concentration are comparably lower than those observed in the main stream of the intensively-cultivated Lanyang-Hsi watershed (0.14–3.5 mg/l; Kao et al., 2004), yet, the lower limit (baseline values) are quite similar to those values reported in non-cultivated tributaries of Lanyang-Hsi. Coefficients of variation at these stations are 0.35, 0.29, and 0.55. Generally, NO$_3$-N concentrations increase with increasing water discharge showing a different pattern to that of Ca and DOC. Station D locates at the highest elevation shows generally lower NO$_3$-N concentrations. While station S at lowest elevation gives higher NO$_3$-N concentrations. At stations, S and K, 3 orders of magnitude of increase in water discharge rate may drive 1 order of magnitude increase in NO$_3$-N concentration; while NO$_3$-N increase at a higher rate per increase of stream flow at station D although observed concentrations are lower.

The concentrations of Si vary in a narrow range (3.52 to 8.74 mg/l) for all stations regardless of fluctuating stream flows. The coefficients of variations of Si at all three stations are also about the same (0.11). The concentrations of Si in samples collected during typhoons were slightly lower compared with normal days. Usually, Si concentration is used to measure chemical weathering and separate proportional contribution from baseflow. Our uniform Si concentration throughout stations and full range of water discharge might indicate a strong weathering condition in Taiwan which may be an interesting issue to study further.
The $\text{SO}_4^-$-$\text{S}$ concentration at D falls between 10.3 and 36.3 mg/l and is consistently higher than those observed at S (1.9–24.5 mg/l) and K (2.3–12.4 mg/l). Coefficients of variation of $\text{SO}_4^-$-$\text{S}$ at stations S, K, and D are 0.54, 0.29, and 0.16, respectively. $\text{SO}_4^-$-$\text{S}$ concentrations decrease when water discharge increases showing a strong dilution effect yet during typhoon period concentrations are relatively stable.

$C$-$Q$ relations of the reference dataset shown in Fig. 4 imply that hydrological responses of constituents are variable. The natural scattering in elemental concentrations under different flow conditions may affect the degree of uncertainty in flux estimation and also influence our selection of estimator if infrequent sampling was applied in the field. Through such a longer term water quality monitoring data, we may push one step forward to evaluate the uncertainty of previous flux estimation reported for SMRs and for future studies as well. Additional information such as the number of samples and the average sampling frequency for each constituent is shown in Table 1, and coefficients of variation are shown in Table 2.

Figure 5 shows the RMSE of flux estimation for five constituents by three estimators at three stations investigated. Because the sizes of the sample pool at each station are different, the x-axis is expressed as the normalized sample size (in %, fraction of sub-group samples in total pool). In most cases, RMSE gradually decreases as sample size increases. Yet, increasing subsample size affects little on RMSE values obtained by using DA method, which apparently gives very stable yet biased estimation. For flow weighted (FW, the middle column in Fig. 5), the decreasing patterns resemble those derived from rating curve (RC) method except $\text{DOC}$ and $\text{NO}_3^-$-$\text{N}$ cases at station D where extremely high errors occur at conditions of small sample size (e.g. 10% of total samples). Such extraordinarily high RMSE values are due to improper subsampling, which miss-catches high flow condition leading to a fault hydrological response (Smart et al., 1999). Thus, extrapolation by using a biased rating curve may enlarge errors significantly at high flow conditions. On the other hand, as increasing number of subsamples RMSE values by RC method decrease more responsively when comparing with the decreasing trend by FW method. Comparing the FW and RC method, we
could find that the errors could be reduced significantly by increasing sampling number as the constitute $C-Q$ plot is highly well-correlated (high $R^2$). The FW and RC method takes the hydrological responses (discharge) into account.

The RMSEs of DA method converge rapidly to a constant value respective to different constituents at different stations. From the mathematical perspective, with increasing numbers of sub-group samples, RMSEs in FW method would be reduced to zero (perfect estimation), whereas in RC method would reduce RMSE to inherent fitting errors due to regressive residuals (reduce to zero after bias correction proposed by Kao et al. (2005)). If the conditions of RMSE lower than 10 for all constituents are required, more than 30% of total samples would be suggested by FW and RC methods. However, for DOC and NO$_3$-N estimation, more samples are necessary because of their high fluctuations.

5 Discussion

5.1 Significance of high-flow sampling

Hydrological condition plays an important role in fluvial material output from the upstream sub-watersheds, in which all constituents’ concentrations vary concomitantly with water discharge rate. From Table 2, the coefficients of variation of stream flow is much greater than that of dissolved materials supporting that water discharge rate in SMRs that fluctuates in several orders of magnitudes is dominating the transport of materials out of the watershed. Consequently, positive relationships between material flux and water runoff can be found usually (Cooper and Watt, 2002) illustrating the flushing nature of material export. This order-of-magnitude positive correlation allows us to establish log-linear rating curves to estimate material export under various water discharge conditions.

High-flow sampling is highly recommended (though not easy) to depict the actual behavior of material flux with stream flow rates. Meanwhile, high flow sampling can avoid
out-of-range extrapolation problems produced by RC method. For example, in the case of DOC and NO$_3$-N at station D, if we only use low-flow samples to construct the rating curve, we would misrepresent the flux behavior and thereby lead to overestimation. The results shown in Figs. 4, 5, and 6 indicate that it is unreliable to use RC method to estimate flux when the samples used to construct the rating curve are less than 10% of total samples. However, the RC method gives a better estimation when adequate samples are obtained. Overall, it is better to use FW method if only few samples are available, with the exception of Ca and SO$_4$-S. These two constituents’ concentrations tend to fluctuate less with increasing water discharge rate; therefore, in these cases, the RC method is better than FW method.

5.2 Effects of C-Q type on flux estimation

Applying the same sampling frequency to different constituent would lead to distinct estimation errors because of the different hydrological responses of constituents. For a discussion of flux estimation uncertainties, the five constituents are classified into four categories: (a) constituent concentration increases with increasing water discharge rate, i.e. NO$_3$-N; (b) concentration decreases with increasing water discharge rate, i.e. Ca and SO$_4$-S; (c) concentration remains stable regardless of water discharge changes, i.e. Si; and (d) no significant relationships between concentration and water discharge rate, i.e. DOC. Figure 6 shows the mean of the relative errors $\bar{\varepsilon}$. When constituent concentration changes belong to condition (a), i.e. NO$_3$-N, a systematic underestimation (except RC method due to extrapolation problem) occurs because of the greater probability of taking samples with lower concentration in low water discharge rate. On the other hand, a systematic overestimation occurs with condition (b). If the C-Q relation conforms to condition (c), i.e. Si, load estimation shows no significant bias either in relative errors or RMSEs among three estimation methods. In most cases, the DOC load estimation is underestimated, except at station K by DA method and at station D by RC method. If we combine Fig. 7 displaying the standard deviation of the relative errors with Fig. 5, it is found that the mean of relative errors of DA method
dominates RMSE. In contrast, the standard deviation of the relative errors by FW and RC methods dominates RMSE. A large extent of standard deviation of relative errors shows that the chosen sampling scheme highly influences the estimation error (either over – or underestimate). The same number of sub-group sample size (sampling numbers) but different set of sub-group samples (sampling schemes) lead to totally different estimation results.

Sampling schemes determine the extent of estimation errors. Thus, the standard deviation of sampled water discharge can be used as an indicator of sampling scheme. Figures 7, 8 and 9 show the sampled $Q$ standard deviation versus relative errors estimated by three estimation methods at three study sites for NO$_3$-N, SO$_4$-S, and Si, respectively. And the red lines represent natural standard deviation of discharge at each station. Fifty subgroup samples are used here to calculate each relative error (each dot) in Figs. 7–9, and there are total 500 different sets of subgroup samples. Most previous study’s suggestions were based on the designed time protocols rather than flow condition. In our opinions, the specific time interval (such as monthly/yearly etc.) suggested by their studies (Thomas and Lewis, 1993, 1995; Kronvang and Bruhn, 1996; Robertson and Roerish, 1999; Moatar and Meybeck, 2005; Quilbe et al., 2006) were not because of appropriate sampling interval but the corresponding well-correlated flow conditions/material concentrations and estimation methods. Thus, we try to correlate the sampled $Q$ standard deviation with estimation error to see how estimation errors change corresponding to sampling schemes. For FW method, the relationship between sampled $Q$ standard deviation and relative error seems to agree with the one between constituent concentration and water discharge rate. With NO$_3$-N as an example, the observed positive relation between NO$_3$-N concentration and water discharge (Fig. 4) also exists between sampled $Q$ standard deviation and the relative error (Fig. 8). For SO$_4$-S, concentration decreases in response to increasing water discharge (Fig. 4) and a similar negative relationship of sampled $Q$ standard deviation versus relative error also exists (Fig. 9) while using FW method. The standard deviation of sampled $Q$ at stations S and K are not continuous in Fig. 8 because of the highly fluctuating water
discharge, which is a characteristic feature of SMRs. High sampled \( Q \) standard deviation means high-flow samples are involved in estimating flux. Samples taken during typhoons, several orders of magnitude changes in water discharge within just a few hours lead to the discontinuity. When the high-flow samples are included in the subgroup samples, a cluster of \( Q \) standard deviation is away from those obtained in normal conditions. For these datasets, usage of RC method can narrow the extent of errors and thus a higher probability for better estimation will be obtained. However, high-flow Si concentrations measured at station S don’t make the estimation better. From the \( C-Q \) relationship, Si concentrations generally remain stable in response to variability of water discharge. However, samples taken during high-flow resulted in underestimation of Si loads, regardless of whether the FW or RC method is used (Fig. 10), due to the low concentrations occurring in the highest water discharge (Fig. 4). Although high-flow samples are suggested with RC method, including such “outlier” contributes nothing for improving estimation accuracy. This result implies that the relative error depends not only on sampled \( Q \) but also on the characteristics in \( C-Q \) plot. Any samples that do not follow the \( C-Q \) relationship of the specific period will bias the estimation results, particularly when event samples are involved. If the observed drop in Si concentration during high-flow is real, then stratified estimation based on magnitude of water discharge rate is further suggested. Overall speaking, if sampled \( Q \) standard deviation matches the natural variability, i.e. standard deviation of daily discharge of concerned periods, FW method is recommended. However, if samples cover the full span of discharge variability, such as selective sampling having samples in the lowest and highest discharge and the sampled \( Q \) standard deviation would be higher than natural variability, the RC method is advocated. For DA method, the extent of relative error does not improve even when high-flow samples are included in the estimation.

### 5.3 Sample size effect on uncertainty

Uncertainty is generally caused by the variability of constituent concentration and the hydrological extent of samples been collected. To achieve lower level of error, more
samples are usually needed to estimate loads if the constituent of interest fluctuates highly with water discharge, such as SO$_4$-S, Ca, and NO$_3$-N. Fewer samples are required and more money and manpower could be saved if the constituent of interest is relatively stable in concentration with flow, such as Si. Besides, higher coefficients of variation of constituent concentration (Table 2) also results in higher RMSE (Fig. 5) as random sampling applied. Flow-stratified sampling (with selective sampling in low- and high-flows properly) coupled with RC method can save manpower by reducing sampling numbers. However, random sampling generates the maxima error according to various sampling schemes. Although the numbers of sample pool of each constituent at every sampling station are slightly different, the trends of RMSE with the increasing percentage of total samples are comparable. If 5% of total samples are taken as a threshold, it is found that DOC with DA method can provide acceptable flux uncertainty of <25%. The RC method is the best flux estimation method for SO$_4$-S, for which an uncertainty of <20% can be reached by just 5% of total samples. Both DA and FW methods are appropriate to apply to NO$_3$-N with 5% of total samples, producing uncertainty of <25%, but the FW method can efficiently narrow the uncertainty if more samples are taken. Ca flux can be appropriately estimated by all three methods, except for the DA method applied to station S. Since Si is the most stable constituent, DA method is good enough to estimate its flux, even with less frequent sampling.

Rank-sum analysis is applied in this study to test if more subsamples produce a statistically significant improvement in flux estimation (Fig. 11). In the plots of rank-sum index versus sample numbers (percentage of total samples), the area between −1.96 to 1.96 on the y-axis (with a 95% confidence interval, the closer to zero, the less significant improvement in flux estimation) represents “statistically insignificant”, which means more samples (increment of x-axis) contribute only limited improvement in flux estimation. On the other hand, y-values beyond the area represent statistically significant improvement. The rank-sum index by DA method is insensitive versus sample number increase (Fig. 11), but the relative errors of estimation converge to a biased level rather than to the best estimation.
To find out the balance between cost (money and manpower used to derive samples) and benefit (accuracy of estimation), the results of rank-sum analysis (Fig. 11) should put together with the RMSE results (Fig. 5). Take DOC by FW method at station S as an example, for sub-group samples more than 50% of total samples can provide more accurate estimation. However, increasing samples from 40% to 50% of total samples only improve 5% RMSE (from around 35% to 30% in Fig. 5). Whether it is worthy or not depends on the decisions of organizations that are responsible for water quality monitoring. They need to make a compromise between estimation errors and cost.

6 Conclusions

Without understanding the background geochemical fluxes, it would be difficult to identify the uncertainty of estimate based on limited samples and how anthropogenic impacts on riverine fluxes. The error distribution of flux estimation in the pristine upstream watersheds due to different sampling numbers and different calculation methods adopted were presented. A random sampling procedure, which provides all the possibility of uncertainties cause by under-sampling, is applied. In light of the relationship between estimation errors and sampling schemes (sampled Q standard deviation), better sampling strategy and estimator are suggested. If sample numbers are sufficient and include high-flow samples, RC method will be recommended; otherwise, FW method is the conservative method. The DA method is recommended only when the target constituent is stable regardless of water discharge changes. Observational and statistical experimental results of small mountainous watersheds discussed here may help constraining better flux values for Oceania Rivers and also save manpower/cost in sampling for many restricted areas and/or ungauged watersheds in future studies and potentially allow researchers to re-examine uncertainties of reported values in the past.

7367
Acknowledgements. This study was supported by grant NSC 95-2116-M-001-001 from the National Science Council of Taiwan. We thank the Water Resources Agency, Taiwan for providing the flow data and AJ Sung, Jean Chen and Jil Chien for collecting and analyzing samples.

References


7370
Table 1. Sampling station, watershed, duration and numbers of each constituent (sampling frequency in the parentheses, unit in days/sample).

<table>
<thead>
<tr>
<th>Station</th>
<th>Watershed</th>
<th>Duration</th>
<th>NO$_3$-N</th>
<th>SO$_4$-S</th>
<th>DOC</th>
<th>Ca</th>
<th>Si</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diyu (S)</td>
<td>Sindian River</td>
<td>22 Apr 2002 ~1 Aug 2005</td>
<td>222*</td>
<td>345</td>
<td>345</td>
<td>345</td>
<td>345</td>
</tr>
<tr>
<td>Dongshe (K)</td>
<td>Keelung River</td>
<td>29 Mar 2002 ~30 Apr 2004</td>
<td>194</td>
<td>195</td>
<td>195</td>
<td>195</td>
<td>195</td>
</tr>
<tr>
<td>Shiouluan (D)</td>
<td>Dahan River</td>
<td>19 Apr 2002 ~15 Apr 2004</td>
<td>189</td>
<td>191</td>
<td>192</td>
<td>191</td>
<td>191</td>
</tr>
</tbody>
</table>
Table 2. Coefficients of variation for all the constituents and stream flow at each of sampling stations.

<table>
<thead>
<tr>
<th>Station</th>
<th>Daily Q</th>
<th>NO$_3$-N</th>
<th>SO$_4$-S</th>
<th>DOC</th>
<th>Ca</th>
<th>Si</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diyu (S)</td>
<td>2.66</td>
<td>0.35</td>
<td>0.54</td>
<td>0.39</td>
<td>0.43</td>
<td>0.11</td>
</tr>
<tr>
<td>Dongshe (K)</td>
<td>2.1</td>
<td>0.29</td>
<td>0.29</td>
<td>0.4</td>
<td>0.24</td>
<td>0.11</td>
</tr>
<tr>
<td>Shiouluan (D)</td>
<td>2.5</td>
<td>0.55</td>
<td>0.16</td>
<td>0.73</td>
<td>0.18</td>
<td>0.11</td>
</tr>
</tbody>
</table>
Fig. 1. Landscape and stream network patterns of Danshuei River and the sampling sites. K, S and D stand, respectively, for Keelung, Sindian and Dahan rivers.
Fig. 2. Comparison of frequency distribution of long term consecutive daily flow (>20 years, solid circles) with flow on sampling days from 2002 to 2005 (open circles).
Fig. 3. Observed discharge (m$^3$/s), concentration of $\text{SO}_4$-$\text{S}$ (mg/l), $\text{NO}_3$-$\text{N}$ (mg/l) and DOC (mg/l) in the Keelung River (at station K) in 2002~2004.
Fig. 4. Relationships between concentration of 5 different constituents (mg/l) and corresponding stream flows (m³/s) on days of sampling at three stations. The open circle at station K represents samples collected during typhoon.

7376
Fig. 5. The root mean square error (RMSE) of estimated constituent loads by different estimation methods over different sample size (from 5 to 95% of total). DA, FW, RC represent direct average, flow-weighted and rating curve methods. The three lines marked by D, K, and S represent the RMSE for different stations.
Fig. 6. The mean of the relative errors of estimated constituent loads by different estimation methods over different sample size. The relative errors are calculated as a percentage of the reference flux. Symbols are the same as those in Fig. 5.
Fig. 7. The standard deviation of the relative errors of estimated constituent loads by different methods over different sample sizes. Symbols are the same as those in Fig. 5.
Fig. 8. The sampled $Q$ standard deviation versus estimation errors for NO$_3$-N. Each triangle represents the relative error of estimated fluxes using 50 sub-samples which are randomly extracted from the sample pool. The y-axis shows the standard deviation of the corresponding stream flow of 50 sub-samples. Red line in each panel represents standard deviation of measured discharge from the sample pool at each station.
Fig. 9. The sampled $Q$ standard deviation versus estimation errors for $\text{SO}_4^-$-S. Other details are the same as those in Fig. 8.
Fig. 10. The sampled $Q$ standard deviation versus estimation errors for Si. Other details are the same as those in Fig. 8.
Fig. 11. Results of rank-sum analysis. The area between two parallel lines in each panel (−1.96 to 1.96 on the y-axis) represents “statistically insignificant” with a 95% confidence interval. Other details are the same as those in Fig. 5.