Hydrology and Earth System Sciences, 3(2), *295–303* (1999) © EGS

Hydrology & Earth System Sciences

Methods for estimating loads transported by rivers

Trevor S. Smart[1], David J. Hirst[2] and David A. Elston*

Biomathematics and Statistics Scotland, Macaulay Land Use Research Institute, Craigiebuckler, Aberdeen, AB15 8QH, UK.
[1] Current address: Pfizer Ltd, Ramsgate Road, Sandwich, Kent CT13 9NJ
[2] Current address: Norsk Regnesentral, PO Box 114 Blindern, N-0314 Olso, Norway
* e-mail for corresponding author: d.elston@bioss.sari.ac.uk

Abstract

Ten methods for estimating the loads of constituents in a river were tested using data from the River Don in North-East Scotland. By treating loads derived from flow and concentration data collected every 2 days as a truth to be predicted, the ten methods were assessed for use when concentration data are collected fortnightly or monthly by sub-sampling from the original data. Estimates of coefficients of variation, bias and mean squared errors of the methods were compared; no method consistently outperformed all others and different methods were appropriate for different constituents. The widely used interpolation methods can be improved upon substantially by modelling the relationship of concentration with flow or seasonality but only if these relationships are strong enough.

Introduction

Estimates of the loads of various constituents transported by a river provide a measure both of the loss of constituents from terrestrial ecosystems and of the environmental impact of the river on the water body into which it flows. Indeed, it is now incumbent on many countries to obtain estimates of loads to ensure compliance with international treaties on the environment (PARCOM, 1988). However, many of the methods used to estimate load either have a high variance or are biased (Walling and Webb, 1985; Ferguson, 1987). Understanding the reliability of the potential estimation procedures, and their interaction with the data collection process, is crucial for selecting the most appropriate method and for interpreting load estimates already obtained.

The instantaneous load being carried by a river at any time, t, is the product of two quantities, namely the flow of the river, $f(t)$, and the concentration of the constituent of interest in the water, $c(t)$. The total load, l, over a period of interest, say from 0 to T, can then be obtained by integration as

$$l = \int_0^T c(t)f(t)\mathrm{d}t.$$

Estimation of this quantity must be based on available data. Whilst flow is generally measured frequently and in many situations can be considered as a continuous record, measurements of concentration are usually less frequent. To formalise this situation, it is assumed that the problem is to estimate load on the basis of N equally spaced observations of flow, f_i, taken at times t_i, $i = 1 \ldots N$, with accompanying observations of concentration, c_i, $i \in S \subseteq \{1 \ldots N\}$, at $n \leq N$ such times. This lack of information about the concentration can result in substantial errors in estimates of the total load. Storms, a rapid snowmelt or a short period of rain in a long dry period can quickly change the rate of flow and the chemical composition of a river. If the sampling is too infrequent, these events may be missed, and, even with fairly frequent sampling, it is difficult to sample a representative number of events.

Choice of method used to estimate load will always have a major bearing on estimates derived from infrequently sampled rivers. Though increasing the sampling frequency will improve the estimation of load, to make the best use of limited numbers of samples other sampling strategies have been considered (Kronvang and Bruhn, 1996; Thomas, 1985). Various methods have been suggested for modelling and, hence, estimating chemical loads (Walling and Webb, 1981 and 1985; Ferguson, 1987; Littlewood, 1995; Kronvang and Bruhn, 1996). These methods fall into two categories: interpolation methods and methods based on empirical models. The key features of these categories of methods are summarised below.

The interpolation methods assume that the data collected are representative of the river at times which are unrecorded. Walling and Webb (1985) discuss five interpolation methods for estimating the load. These five methods are applications of standard statistical sampling theory, and provide estimates of load as follows:

$$\hat{l} = K\left\{\sum_{i\in S} c_i / n\right\}\left\{\sum_{i\in S} f_i / n\right\} \qquad \text{W1}$$

$$\hat{l} = K\left\{\sum_{i\in S} c_i f_i / n\right\} \qquad \text{W2}$$

$$\hat{l} = K\left\{\sum_{i\in S} c_i F_i / n\right\} \qquad \text{W3}$$

$$\hat{l} = Kf_{\cdot}\left\{\sum_{i\in S} c_i / n\right\} \qquad \text{W4}$$

$$\hat{l} = Kf_{\cdot}\left\{\sum_{i\in S} c_i f_i / n\right\}\Bigg/\left\{\sum_{i\in S} f_i / n\right\} \qquad \text{W5}$$

where K is a constant which adjusts for units of measurement, $f_{\cdot}$ is the mean discharge over all times t_i, $i = 1 \ldots N$ and F_i is the mean discharge between time t_i and the next time at which concentration was measured. Methods W1 and W2 do not use the entire flow record; W1 assumes concentration is independent of flow and gives biased estimates if this assumption is violated, whilst W2 is unbiased but has higher variance than W1. Methods W4 and W5 are generalisations of W1 and W2 to use the complete flow record, although other generalisations are also used (Young *et al.*, 1988). Method W3 lies between W4 and W5 in both bias and variance.

Empirical models, referred to by some authors as extrapolation methods* or rating curves, are usually based on the linear relationship between log(concentration) and log(flow). The observations are used to estimate the parameters in a model and the fitted model then predicts the concentrations using flow data from other times. This procedure assumes that the relationship between the concentration and the other variables is representative of their relationship throughout the period of interest. In these empirical models, problems can arise if the model is extrapolated beyond the range of the data used to fit the model, as may be required if no concentrations were measured at extreme events. Additionally, bias can arise if the models describe the concentration-flow relationship incorrectly and the distribution of flows used to estimate parameters differs from the distribution of flows to which the model is applied for load estimation. Elaborations on the basic theme have been made to incorporate features such as seasonal effects and stratification according to flow (Walling and Webb, 1981; Kronvang and Bruhn, 1996).

Assessments of methods for estimating loads have been carried out using simulated data (Littlewood, 1995; Thomas, 1985), or using data from intensively sampled rivers (Walling and Webb, 1981, 1982 and 1985; Kronvang and Bruhn, 1996), or both (Ferguson, 1987). The simulation route has the advantage that a wide range of situations can be catered for, but it will tend to favour methods based on models used to generate the data. Conversely, the use of case studies provides the most realistic test but, by their very nature, the number of such studies is limited and care is required in understanding the reasons for results found. What does emerge from these studies is that the choice of method depends on the selection criterion and the characteristics of river and constituent used or simulated.

In this paper, an extensive set of flow and concentration measurements was used to test the ability of methods W4, W5 and a range of empirical models to estimate total loads. Loads are estimated for constituents whose correlations with flow vary from being very weak to moderately strong. By repeated random sub-sampling from the data available, subsets were obtained whose sampling intensity for concentrations more closely matches that found in practice. By estimating load for each subset, the differences in bias and variability between the methods has been demonstrated, to provide advice on which methods will be most appropriate in other circumstances.

Method

THE DATA

Measurements of flow and concentration of various constituents were taken for the River Don every two days, for 11.4 years from 1 August 1976 until 23 December 1987. The recording site lay at Parkhill Bridge, just above the tidal limit but within the boundary of Aberdeen City. The 1400 km^2 catchment area of this river in North East Scotland includes part of the Cairngorm Mountains and is composed mainly of areas of semi-natural moorland, forestry and agriculture. Time series of nitrate, chloride, phosphate, suspended solids and flow at the time of sampling are shown in Figs. 1a to 1e. The nitrate and phosphate concentrations are reported on an elemental basis for nitrogen (nitrate-N) and phosphorous (phosphate-P) respectively. The graph of the flow series shows that flow had a dynamic range and that many storms were recorded. The four constituents, whose log-transformed concentrations are plotted against log(flow) in Fig. 2, were chosen for their very different characteristics. Nitrate had a positive correlation with flow and a very marked seasonal pattern. Chloride had no clear relationship with flow; it also had the lowest coefficient of variation. Phosphate had a negative correlation with flow, whilst the majority of the load of suspended solids was transported at very high concentrations at times of high flow.

* The term empirical modelling rather than extrapolation methods has been used. The problem of extrapolating beyond the range of the data used to form the model is just one specific problem of these empirical models and is an issue that needs to be addressed separately.

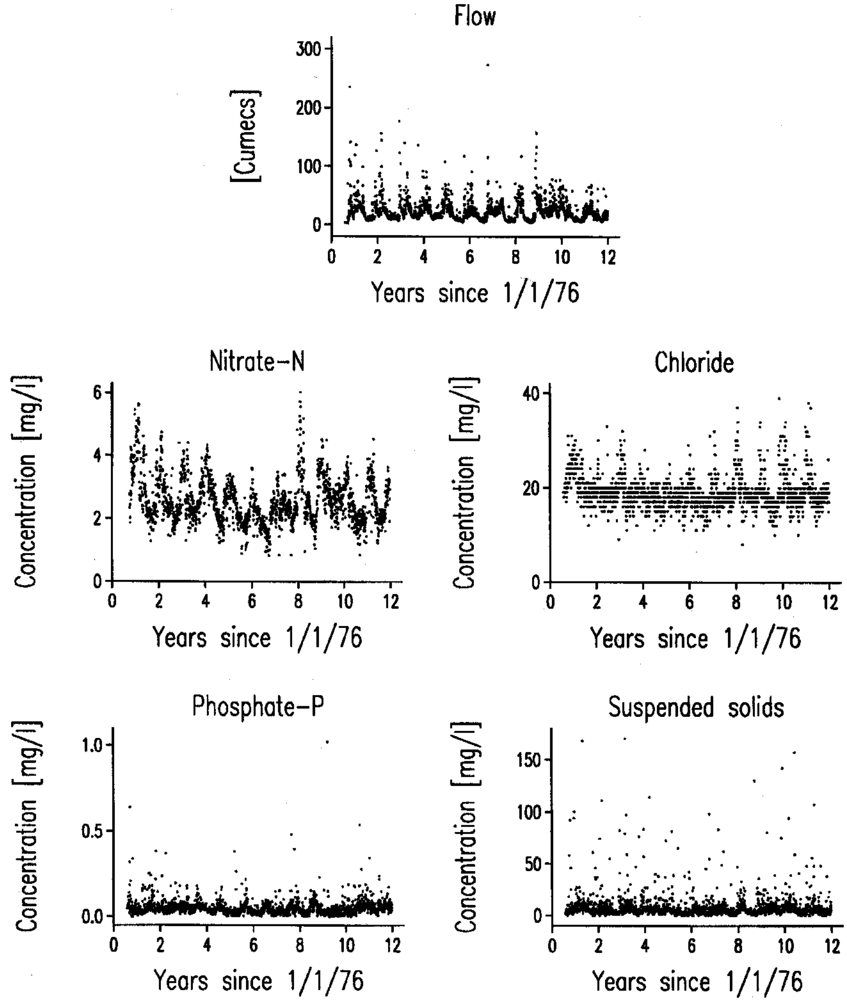

Fig. 1. *Time series of a) flow, b) nitrate-N, c) chloride, d) phosphate-P and e) suspended solids.*

Procedure for comparing methods of load estimation

This paper assesses different methods for estimating load under the assumption that the flow was measured every two days but that the concentrations were measured only every 2 or 4 weeks. This reflects the common situation where flow is measured at regular time intervals and at a much greater frequency than for concentrations. Estimated loads are then compared with the corresponding value calculated using method W2 with all flow and concentration data, which is treated as a truth to be predicted.

To compare the sampling properties of the methods of load estimation, 1000 sequences of data were generated as follows. The recorded concentrations and flows were split into 14- or 28-day blocks. One sample was selected at random from each block. From these sub-samples of concentrations, but using the flow from every second day, estimates of the total loads were found using each different method. This gave a total of 1000 estimates for each chemical for each method from which three sampling properties, namely coefficient of variation, bias and root mean squared error were obtained and expressed as a percentage of total load over the entire period. Use of 1000 simulations allowed sufficiently accurate estimation of these summary statistics to show up even quite small differences, but did not enable determination of whether the differences were a general result, or were due solely to the particular period for which data were available.

To derive estimates of variability which included a component for between-year variation, a jack-knife was performed by groups (Miller, 1974); the entire procedure described above was repeated with each of the 11 full years of data dropped in turn. Variation between pseudo-values across years was used to derive the standard errors presented, whilst standard errors of difference between methods were derived from differences between pseudo-values corresponding to the same years. This approach allowed inferences to be drawn beyond the particular years in question, although there are no empirical statements

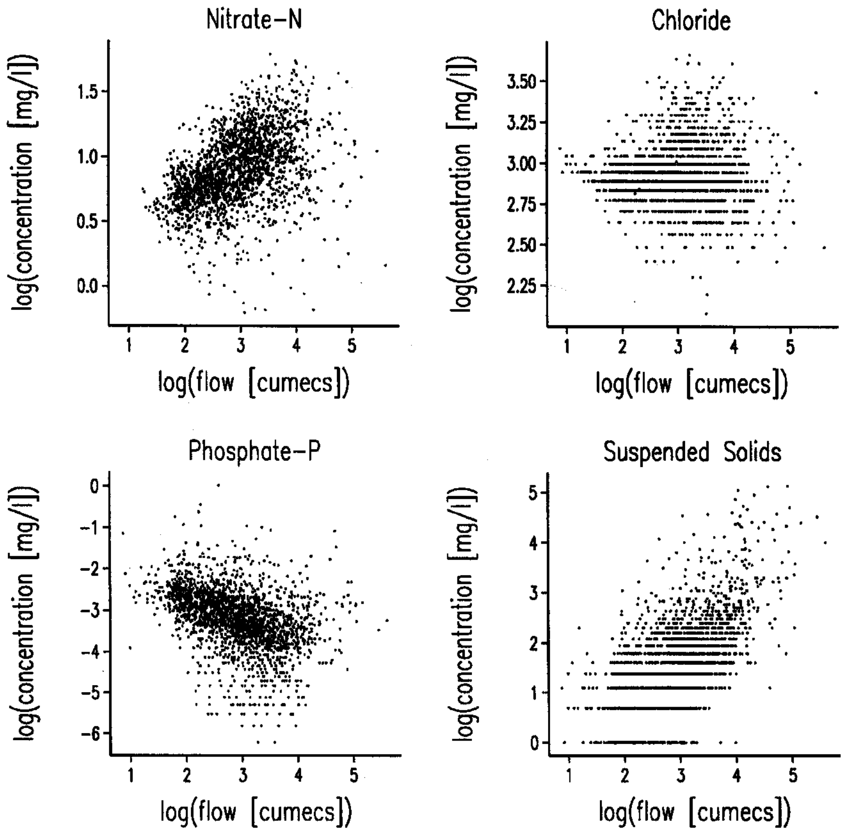

Fig. 2. *Graphs showing the relationship between flow and a) nitrate-N, b) chloride, c) phosphate-P and d) suspended solids.*

which can be made about the extension of these results to other rivers.

Development of empirical models

The two most promising of the interpolation methods, W4 and W5, were compared with estimates derived from some simple log linear models,

$$\log(c_i) = a_0 + a_1 x_{1i} + a_2 x_{2i} + \ldots + a_m x_{mi} + \varepsilon_i.$$

The residual term, ε_i, was assumed to come from a normal distribution with expectation zero and variance σ^2. The coefficients, a_k, $k = 1 \ldots m$, were estimated from the sample data only and used to predict the log concentration at other times. Whilst the regression models provide unbiased estimates of $\log(c_i)$, $i = 1 \ldots N$, it was estimates of c_i, $i \notin S$, that were needed to estimate the load. Naive back-transformation of the log concentrations resulted in a negative bias in the estimated concentrations (Ferguson, 1986). Therefore, the estimates of concentration were rescaled, assuming the lognormal model, by

$$\hat{c}_j = \exp(\hat{\sigma}^2 / 2)\exp(\log\hat{}(c_j)),$$

where $\hat{\sigma}^2$ is the residual mean square of the model and $\log\hat{}(c_j)$ is the estimated log concentration.

The simplest models included a regression on the log of the flow, i.e. $m = 1$, $x_{1i} = \log(f_i)$, or seasonality described by a sine curve, i.e. $m = 2$, $x_{1i} = \sin(2\pi t_i/365)$, $x_{2i} = \cos(2\pi t_i/365)$ with time, t_i, recorded in days. Further models tested are described below. Many other candidates could have been tested: non-linear relationships between log(concentration) and log(flow) or a less constrained seasonal pattern (Miller and Hirst, 1998) provide alternative ways of handling the systematic parts of the model, whilst correlations in the residuals can be handled using time series models (Gurnell and Fenn, 1984) or one-dimensional kriging. However, it seems likely that these more complicated models would require more frequent data than generally available, and so were not tested.

Taking account of the order of the observations

The correlation between concentrations measured over short time periods can be large but, when using only $\log(f_i)$ as the explanatory variable, no account is taken of the order of the measurements. One possibility is to smooth the sequence of estimated concentrations (or log concentrations) prior to estimating the load. This will also remove some of the extreme values caused by extrapolation. Smoothing used a weighted, moving average of the estimated log concentrations, thus:

$$\log\tilde{g}(c_i) = 0.25 \times \log\hat{g}(c_{i-1}) + 0.5 \times \log\hat{g}(c_i) + 0.25 \times \log\hat{g}(c_{i+1})$$

Another possibility is to use some measure of the previous flow as a regressor, partly to account for hysteresis, which can have an important effect on the concentration (Hirst, 1992). The mean log(flow) for the eight days prior to day t_i, g_i, calculated as

$$g_i = (\log(f_{i-4}) + \log(f_{i-3}) + \log(f_{i-2}) + \log(f_{i-1})) / 4$$

was therefore included in some models. Calculating g_i in this way was somewhat arbitrary and other values could clearly be used, although the chosen method was found to have some benefits.

Reducing extrapolation

Any empirical model will have only limited predictive value outside the range of the data used to determine the functional form and to estimate parameters. This is particularly a problem when trying to predict concentrations at high flows. Moreover, the problem is worse if a variable correlated with flow, such as seasonality, is included in the model, as this may increase the variance of the estimate of the flow coefficient. The approach adopted to reduce the impact of extrapolation was to treat as cut-off values the maximum, f_{max}, and minimum, f_{min}, of the flows corresponding to measured concentrations. For prediction purposes, any flows above f_{max} were replaced by f_{max}, and any flows below f_{min} were replaced by f_{min}. Such an approach introduces some bias into the model but it will also reduce the extra variance caused by extrapolation.

Methods to be compared

The ten methods compared fall into three categories: interpolation methods, log linear models, and log linear models with additional smoothing or cut-offs. Units of measurement are time, d, flow, m^3s^{-1}, and concentration, mgl^{-1}. Taking $K = 0.03156$ gives estimates, $\hat{l}$, of mean annual load, $kT\,y^{-1}$.

INTERPOLATION METHODS

1) $$\hat{l} = K\bar{f}\left\{\sum_{i\in s} c_i / n\right\}$$ W4

2) $$\hat{l} = K\bar{f}\left\{\sum_{i\in s} c_i f_i / n\right\} \Big/ \left\{\sum_{i\in s} f_i / n\right\}$$ W5

EMPIRICAL MODELS

Estimation was based on the equation

$$\hat{l} = K\left\{\sum_{i=1}^{N} \hat{c}_i f_i / N\right\}$$

where $\hat{c}_i$ is the measured concentration where this exists and the estimated concentration with bias correction elsewhere.

Log-linear models

3) Flow only:
$\log(c_i) = a_0 + a_1 \log(f_i) + \varepsilon_i$, M1
4) Seasonality fitted as a sine curve:
$\log(c_i) = a_0 + a_1 \sin(2\pi t_i/365) + a_2 \cos(2\pi t_i/365) + \varepsilon_i$, M2
5) Flow plus previous flow:
$\log(c_i) = a_0 + a_1 \log(f_i) + a_2 g_i + \varepsilon_i$, M3
6) Flow, previous flow, seasonality and linear time trend: $\log(c_i) = a_0 + a_1 \log(f_i) + a_2 g_i + a_3 \sin(2\pi t_i/365) + a_4 \cos(2\pi t_i/365) + a_5 t_i + \varepsilon_i$ M4

Log-linear models with additional smoothing or cut-offs

7) M1 with smoothing over time of the estimated log concentrations. SM1
8) M1 with cut-offs for flow for extrapolation. CM1
9) M4 with smoothing over time of the estimated log concentrations. SM4
10) M4 with cut-offs for flow for extrapolation. CM4

Results

Estimates of bias and sampling variation, presented as coefficients of variation, CV, relative to the corresponding true loads, are given in Table 1a and b, for fortnightly and monthly sampling of the water chemistry respectively. The positive correlations between methods within years meant that the standard errors of differences were about 40% of those which would have been calculated ignoring this correlation, although for suspended solids the added variability led to reduced correlation between methods. Note also that standard errors of difference (sed) for bias depended much less on sampling frequency than did those for CV.

The bias and variance of the estimated loads have been combined to form mean squared errors, which are presented after dividing by the true load as relative root mean squared error (RRMSE) (Table 2a and b). Again, the correlation between methods within years was least for suspended solids.

The relative merits of the methods of estimation are summarised by element below. All of these results were remarkably consistent across sampling frequencies. Actual loads for nitrate-N, chloride, phosphate-P and suspended solids calculated using the 48-hour sample data were 2.054, 14.396, 0.0339 and 10.86 kTy^{-1} respectively.

Table 1. Coefficient of variation (CV) and bias, with standard errors from the jack-knife in parentheses, presented as a percentage of the actual total load transported by the river, for each of the ten methods and the four chemicals using (a) fortnightly sampling and (b) monthly sampling. Also given are the mean pairwise standard errors of differences, together with the mean ratios of paired standard errors of differences to standard errors of differences calculated ignoring the pairing.

a Concentration measured once per fortnight

Method	Nitrate		Chloride		Phosphate		Suspended Solids	
	CV(%)	Bias(%)	CV(%)	Bias(%)	CV(%)	Bias(%)	CV(%)	Bias(%)
W4	0.78(0.08)	–6.42(1.86)	0.72(0.06)	–0.92(0.65)	5.25(0.99)	16.96(3.05)	4.94(0.50)	–44.83(2.87)
W5	1.82(0.27)	1.03(0.33)	1.42(0.21)	–0.09(0.22)	6.46(0.75)	0.38(0.43)	17.97(1.81)	–1.27(2.33)
M1	1.82(0.30)	2.53(0.99)	1.11(0.12)	–0.14(0.34)	4.62(0.44)	–3.83(1.26)	7.15(0.70)	–16.86(2.79)
M2	0.95(0.11)	–0.73(1.21)	0.86(0.07)	1.28(0.42)	4.17(0.60)	8.54(3.50)	5.74(0.47)	–37.11(2.78)
M3	1.67(0.29)	0.22(0.75)	1.20(0.13)	–0.88(0.34)	4.64(0.48)	–2.69(1.26)	17.03(3.34)	14.20(4.65)
M4	1.52(0.25)	–0.35(0.73)	1.10(0.10)	–1.10(0.30)	4.70(0.46)	–2.01(1.16)	15.21(2.73)	9.23(4.08)
SM1	1.65(0.26)	1.51(0.91)	1.07(0.11)	–0.21(0.31)	4.60(0.44)	–2.72(1.25)	6.45(0.56)	–22.59(3.19)
CM1	1.76(0.26)	2.37(0.89)	1.11(0.11)	–0.15(0.34)	4.61(0.43)	–3.71(1.19)	7.14(0.65)	–18.16(3.20)
SM4	1.38(0.22)	–0.04(0.79)	1.02(0.09)	–0.90(0.31)	4.39(0.43)	–2.10(1.15)	5.70(0.69)	–20.85(2.33)
CM4	1.50(0.24)	–0.32(0.74)	1.09(0.10)	–1.07(0.30)	4.67(0.46)	–1.98(1.15)	12.73(3.45)	2.72(6.61)
Mean sed	0.17	0.75	0.07	0.33	0.38	1.64	1.88	3.86
Mean sed ratio	0.52	0.54	0.44	0.62	0.44	0.60	0.74	0.72

b Concentration measured once per month

Method	Nitrate		Chloride		Phosphate		Suspended Solids	
	CV(%)	Bias(%)	CV(%)	Bias(%)	CV(%)	Bias(%)	CV(%)	Bias(%)
W4	1.32(0.14)	–6.39(1.79)	1.09(0.10)	–0.98(0.60)	7.95(1.63)	16.47(3.23)	7.10(0.52)	–44.34 (2.76)
W5	2.92(0.35)	0.97(0.48)	2.09(0.19)	–0.15(0.17)	9.31(1.20)	–0.12(0.96)	24.98(2.18)	0.04 (1.93)
M1	2.95(0.49)	2.75(1.02)	1.69(0.13)	–0.19(0.40)	6.58(0.57)	–4.17(1.26)	9.84(1.12)	–17.55 (2.30)
M2	1.57(0.19)	–0.85(1.20)	1.31(0.09)	1.33(0.39)	6.33(0.95)	9.33(3.97)	7.10(0.46)	–39.42 (2.68)
M3	2.76(0.40)	0.34(0.79)	1.82(0.14)	–0.92(0.37)	6.84(0.50)	–2.78(1.24)	25.80(7.83)	17.51 (7.64)
M4	2.50(0.34)	–0.31(0.67)	1.68(0.11)	–1.19(0.27)	6.96(0.54)	–1.83(1.25)	23.46(6.64)	12.17 (6.57)
SM1	2.69(0.42)	1.64(0.90)	1.62(0.12)	–0.27(0.36)	6.58(0.59)	–2.99(1.24)	8.43(0.80)	–23.84 (2.89)
CM1	2.79(0.38)	2.41(0.82)	1.67(0.12)	–0.21(0.38)	6.54(0.58)	–3.91(1.14)	9.74(1.09)	–19.86 (3.43)
SM4	2.30(0.30)	–0.02(0.73)	1.55(0.10)	–0.98(0.29)	6.42(0.48)	–2.07(1.24)	8.09(1.46)	–21.89 (2.13)
CM4	2.44(0.29)	–0.24(0.70)	1.65(0.11)	–1.12(0.32)	6.85(0.50)	–1.77(1.24)	18.29(7.10)	0.73(11.66)
Mean sed	0.25	0.82	0.10	0.36	0.69	1.84	4.16	6.12
Mean sed ratio	0.53	0.59	0.56	0.69	0.55	0.63	0.79	0.84

Nitrate

M2, the method which modelled the effect of seasonality alone, gave the best estimates of load for nitrate. For both sampling frequencies, M2 had CVs that were significantly ($p<0.01$) lower than those for all other methods except W4. Whilst W4 was the least variable method, the large negative bias led to W4 having the largest RRMSE. The three methods based on M1, involving flow alone, had significant positive bias ($p<0.05$) relative to all other methods except W5. The use of smoothing improved methods M1 and M4 with respect to both bias and variance, whereas the introduction of a cut-off was considerably less effective.

Chloride

Estimates of load were generally better for chloride than for any other chemical, with most values of RRMSE lying between 1 and 2%. The methods based on M1, using the simple regression of log(concentration) on log(flow), all had little bias, whereas M3 and the methods based on M4 all showed negative bias. The least variable methods again were W4 followed by M2, the latter being the only method to show positive bias. This bias in M2 gave it one of the greatest values of RRMSE, whereas the smallest values of RRMSE occurred for W4 along with M1 and its adaptations.

Phosphate

Phosphate was negatively correlated with flow, hence the methods which did not include flow, W4 and M2, both had substantial positive bias. W5 had the minimum bias but a CV that was significantly ($p<0.05$) greater than all other methods except W4. After discounting these three methods, there was less to choose between the remainder. M3 and M4 had significantly less bias than M1 ($p<0.05$), but similar CVs. This, combined with the small but significant benefits of cut-off and smoothing, led to SM4 having the smallest RRMSE, followed by CM4, for both sampling intervals.

Table 2. Relative root mean square error, with standard errors in parentheses, presented as a percentage of the actual total load transported by the river, for each of the ten methods and the four chemicals using (a) fortnightly and (b) monthly sampling. Also given are the mean pair-wise standard errors of differences, together with the mean ratios of paired standard errors of differences to standard errors of differences calculated ignoring the pairing.

a Concentration measured once per fortnight

Method	Nitrate	Chloride	Phosphate	Suspended Solids
W4	6.46(1.85)	1.17(0.49)	17.74(3.02)	45.10(2.80)
W5	2.08(0.34)	1.42(0.21)	6.46(0.75)	18.11(1.80)
M1	3.11(0.94)	1.12(0.11)	5.99(0.95)	18.30(2.46)
M2	1.20(0.50)	1.54(0.35)	9.49(3.12)	37.55(2.74)
M3	1.68(0.41)	1.49(0.21)	5.35(0.73)	22.10(5.08)
M4	1.56(0.25)	1.55(0.25)	5.10(0.58)	17.73(4.12)
SM1	2.23(0.76)	1.09(0.11)	5.33(0.81)	23.49(3.00)
CM1	2.95(0.82)	1.11(0.10)	5.90(0.89)	19.51(2.90)
SM4	1.38(0.25)	1.35(0.25)	4.86(0.61)	21.61(2.23)
CM4	1.54(0.23)	1.52(0.24)	5.06(0.57)	13.01(5.28)
Mean sed	0.83	0.26	1.37	3.53
Mean sed ratio	0.78	0.69	0.58	0.71

b Concentration measured once per month

Method	Nitrate	Chloride	Phosphate	Suspended Solids
W4	6.53(1.76)	1.47(0.36)	18.26(3.46)	44.90 (2.69)
W5	3.08(0.45)	2.10(0.19)	9.29(1.21)	25.00 (2.16)
M1	4.02(1.00)	1.70(0.12)	7.78(0.90)	20.13 (1.88)
M2	1.78(0.40)	1.87(0.30)	11.26(3.60)	40.06 (2.64)
M3	2.77(0.52)	2.04(0.14)	7.38(0.72)	31.05(10.48)
M4	2.51(0.33)	2.06(0.14)	7.19(0.64)	26.31 (8.77)
SM1	3.14(0.76)	1.64(0.11)	7.21(0.81)	25.29 (2.63)
CM1	3.68(0.77)	1.68(0.12)	7.61(0.85)	22.13 (3.12)
SM4	2.29(0.33)	1.84(0.14)	6.73(0.64)	23.34 (1.81)
CM4	2.44(0.29)	2.00(0.18)	7.07(0.60)	18.23 (9.92)
Mean sed	0.77	0.23	1.57	5.89
Mean sed ratio	0.71	0.83	0.58	0.75

Suspended Solids

Suspended solids provided the hardest loads to estimate, with most RRMSEs around 20% or more. No method performed consistently well for bias and variance: those methods which had the least bias had the highest values of CV, and vice-versa. Since the majority of suspended solids were transported when the flow was at an extreme high, W4 and M2 were highly negatively biased, to the extent of around –40%. Conversely, methods M3 and M4 both showed positive bias. The introduction of smoothing and a cut-off had little effect on M1 but had striking effects on M4. Smoothing on M4 caused a highly significant ($p<0.001$) change in bias from positive to negative, with a compensating highly significant ($p<0.005$) decrease in CV by a factor of nearly 3. The most variable methods were W5, M3 and M4. The compensation between bias and variance in estimating suspended solids was such that, for RRMSE, after the poorest methods (W4 and M2) had been discarded, there was little to choose between the remainder.

Discussion

The aim of this paper has been to use the frequently sampled chemical concentrations in the River Don to provide a truth against which to assess ten methods of estimating loads. The validity of extending the comparative sampling properties from this simulation to estimates of mean annual load depends primarily on two weak assumptions, namely that the mean load at observation times is well estimated by W2 and that the data span a representative range of conditions experienced by the river during such a long time period. Note that, in absolute terms, the sub-sampling procedure will have led to under-estimates of variability in load estimates since any two sub-samples will have shared some concentration data. However, the investigations suggested the underestimation was quite small and the comparison of methods remains valid since this underestimation affects all methods equally. The use of standard errors of difference based on the use of jack-knifing by years allowed inferences to be drawn that are not restricted to the years for which data were available, although clearly the empirical results are still restricted to the River Don.

The results for estimating loads on the River Don can be summarised as follows.

1 Loads of some elements are much easier to estimate than others. With fortnightly concentration data, chloride can be predicted with a RRMSE of less than 2%, whereas for suspended solids it is nearer 20%. This, of course, reflects greater inherent uncertainty in the estimation of loads of suspended solids and may call into question the value of any method using concentration data collected so infrequently.

2 If there is a negative correlation between concentration and flow, ignoring this will give a positive bias (e.g.

phosphate). If the correlation is positive the bias will be negative (suspended solids). Whether or not to model this correlation depends on the balance between increasing the variance by adding more parameters to the model, and reducing the bias. If there is only a weak relationship between concentration and flow (e.g. chloride), the biased method W4 may be best.

3 If there is a clear seasonal pattern (e.g. nitrate), this can be modelled using a sine curve to improve prediction.

4 In some cases (e.g. phosphate), using a measure of previous flow may improve prediction.

5 Smoothing the estimated concentrations improves the predictions for all chemicals except suspended solids. The smoothing exploits the time series nature of the data, reducing the variance but increasing the bias. It fails with suspended solids because most of the load is due to very high concentrations on a few occasions.

6 Reducing the effect of extrapolation by cutting off the high and low flows gives a small improvement. The effect is similar to that for smoothing described above, i.e. reducing variance but increasing bias, and its benefits are particularly noticeable for suspended solids.

7 The best model is different in each case, but large improvements are possible over the 'standard' method W5.

8 Changing the frequency with which water quality data are available by a factor of two does little to change the relative sampling properties of the methods assessed.

In this paper, estimated loads have been compared with a 'true' load but, clearly, this is not usually possible. Indeed, when only limited concentration data are available, a systematic study based on that data alone may prove poor at discriminating between the methods available. Instead, it may be better to reach a decision about which method to use based on knowledge about the river, the constituent, the data available, results of case studies such as this and results based on simulations.

The criterion used for choosing the best method for load estimation will depend on the trade-off between bias and variance. For a method based on an empirical model, the variability of estimated load during a short period will be greatly reduced by fitting the model to data from both before and after that period, although this can cause bias if the water quality is subject to long-term trends. Choosing the length of data sequence for parameter estimation is, thus, a direct choice between bias and variance. Where the primary interest is to estimate the change in load over time, the method with the minimum variance will be optimal provided it can be assumed that the bias will remain constant. Conversely, if the total load from several rivers is to be estimated, it may be more important to have an unbiased estimate for each river as the CV for the combined load will be less than the CV for the individual loads.

Conclusions

This case study reinforces the view that no single method is consistently best for estimating total loads. Each individual case requires separate consideration. The best method will depend on:

- the element for which the load is to be estimated;
- the river in which the measurements were taken;
- the number and frequency of samples;
- the flow/concentration relationship in the observed data;
- the difference in distribution of the flow when both flow and concentration were observed and when only flow was observed;
- the objective behind estimating the load.

Acknowledgement

Thanks are due to the Scottish Environmental Protection Agency for the use of their data and to Dr. A. Edwards, MLURI, for comments on the methods used and draft texts. Estimation of errors using the jack-knife by years was suggested by a referee. This work was funded by the Scottish Office Agriculture, Environment and Fisheries Department.

References

Ferguson, R.I., 1986. River loads underestimated by rating curves. *Wat. Resour. Res.*, **22**, 74–76.

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

Gurnell, A.M. and Fenn, C.R., 1984. Box-Jenkins transfer function models applied to suspended sediment concentration-discharge relationships in a proglacial stream. *Arctic Alpine Res.*, **16**, 93–106.

Hirst, D., 1992. A new technique for the analysis of continuously monitored water-quality data. *J. Hydrol.*, **134**, 95–102.

Kronvang, B. and Bruhn, A.J., 1996. Choice of sampling strategy and estimation method for calculating nitrogen and phosphorus transport in small lowland streams. *Hydrol. Processes*, **10**, 1483–1501.

Littlewood, I.G., 1995. Hydrological regimes, sampling strategies, and assessment of errors in mass load estimates for United Kingdom rivers. *Environ. Internat.*, **21**, 211–220.

Miller, J.D. and Hirst, D.J., 1998. Trends in concentrations of solutes in an upland catchment in Scotland. *Sci. Total Environ.*, **216**, 77–88.

Miller, R.G., 1974. The jackknife—a review. *Biometrika*, **61**, 1–15.

PARCOM, 1988. Principles of the comprehensive study on riverine inputs. *Annex 9, Tenth Meeting of the Paris Commission*, Lisbon, 15–17 June 1988.

Thomas, R.B., 1985. Estimating total suspended sediment yield with probability sampling. *Wat. Resour. Res.*, **21**, 1381–1388.

Walling, D.E. and Webb, B.W., 1981. The reliability of suspended sediment load data. In: Erosion and Sediment Transport Measurement, *IAHS Publ. No. 133*, 177–194.

Walling, D.E. and Webb, B.W., 1982. The design of sampling programs for studying catchment nutrient dynamics. *Proceedings of an International Symposium on Hydrological Research Basins and their Use in Water Resources Planning*, 21–23 September 1982, Bern, 3, 747–758.

Walling, D.E. and Webb, B.W., 1985. Estimating the discharge of contaminants to coastal waters by rivers: some cautionary comments. *Marine Pollut. Bull.*, 16, 488–492.

Young, T.C., DePinto, J.V. and Heidtke, T.M., 1988. Factors affecting the efficiency of some estimators of fluvial total phosphorous load. *Wat. Resour. Res.*, **24**, 1535–1540.