Comparison of load estimation methods and their associated error

Marsh, N [1], Waters, D [2]

[1]*eWater Cooperative Research Centre, CSIRO Land and Water, 120, Meiers Road, Indooroopilly, 4068.*

[2]*eWater Cooperative Research Centre, Qld Department of Natural Resources and Water, Toowoomba, 4350*
Email: nick.marsh@csiro.au

Extended abstract

The calibration or validation of constituent generation models generally requires the calculation of mass loads from measured concentration data. The determination of load is simply a function of concentration and flow. However concentration data is usually inadequate to perform a direct integration of flow and concentration. As a consequence, many alternative load estimation methods have been suggested to provide a mechanism to calculate loads. The choice of load estimation method has a large bearing on the estimated load of up to an order of magnitude.

Ignoring the errors associated with field sampling and laboratory techniques, the error of any given load estimation is a combination of load estimation method applied to the data and the sampling regime used. For model calibration it is valuable to know if the measured 'true load' against which one is calibrating is within an order of magnitude of the likely true load.

To date there has been no consistent use of load estimation methods and no quantification of the error in load estimation. Previous studies also focused on long term (average annual) load estimation whilst this study focuses on load estimation for shorter duration events (1-50 days). We have applied 34 alternative load estimation methods to 31 storm events where flow and total suspended sediment (TSS) concentration data were available for tropical and subtropical rivers in Queensland. It was a prerequisite that events had: a minimum of 10 samples collected, some on the rise, peak and fall of the hydrograph.

To assess the interaction between sampling regime and load estimation method, we used linear interpolation values between the sampled values to generate high frequency (1500 equal duration time-steps) 'synthetic' data sets. The 31 synthesised high sample-frequency storm events were then resampled using 250 different sampling methodologies.

A Monte Carlo analysis was then undertaken for the 250 different sampling methodologies, whereby each method was applied 50 times to each high frequency data set. We then used each of these data sets to determine the estimated load for the 34 alternative load estimation methods.

The most robust sampling method was defined as that which gave the lowest root mean squared error (RMSE) values across all load estimation methods and events. The 34 load estimation methods can be split into three categories; regression, ratio and averaging. Our results show that regression techniques are the poorest performing with ratio methods being consistently more robust predictors of true load for sample sizes ranging from 3-20 distributed across the hydrograph.

This work will give water quality monitors and modellers a greater understanding of the most robust load estimation method for any given sampling methodology as well as providing confidence intervals for each load estimation and sampling method combinations.

Keywords: *load estimation, sediment, water quality*

1. INTRODUCTION

To quantify load estimates for diffuse and point sources of sediment and nutrients from current land management practices, there has been a move in recent years to measure, model and regulate on the basis of loads of sediment and nutrients in streams (Keyes and Radcliffe, (2002). In order to quantify base loads, or to calibrate or validate a catchment model used to predict load generation, we require a robust, accurate and repeatable method of quantifying loads.

The need to combine stream flow data collected at high frequency with different constituent concentration sampling regimes used within and between rivers has given rise to a large number of load estimation methods. Ignoring the variation in concentration values due to field sampling and laboratory techniques, the error of any given load estimation is a combination of the sampling regime used and the load estimation method applied to the data. To establish future constituent load reduction targets, the error of the original and any future load estimates must be known, otherwise the successful achievement or otherwise of a target load reduction may simply be an artifact of the unquantified error in load estimation.

The variability in load estimates for a given flow and concentration data set are well described (e.g. Walling and Webb, 1981), and in an attempt to provide guidance on appropriate methods several authors have conducted comparative analysis across load estimation methods and sampling techniques, (Kronvang and Bruhn, 1996; Richard and Holloway, 1987; Ferguson, 1987; Preston, *et al*., 1989; Webb *et al.*, 1997; Horowitz, 2003). The general approach of these authors is to use the highest sample frequency data set available (usually monthly) to compare load estimation methods for determining annual load, sometimes with the data set decimated to represent alternative sampling regimes. Monthly sampling may not capture the period when highest loads may occur. As pointed out by Quilbe *et al.,* (2006), how high-flow events are sampled plays a key role in load estimator performance since these events often transport the bulk of the annual load in a short-time period". Many tropical and subtropical rivers are effectively ephemeral in terms of sediment delivery, with load delivery dominated by infrequent short duration runoff events. Walling and Webb, (1981) showed that 80% of the sediment load was delivered in 3% of the time for a temporal stream in the UK.

A review of existing literature found that many of the studies were method and or site specific. Whilst individual studies are very useful in describing site and data set specific load estimation methods, the broader application of the studies was not always clear. For example studies by (Dolan *et al*., 1981, Preston *et al*., 1989, Young and DePinto, 1988, Richard and Holloway, 1987) all showed the Beale ratio estimator to be a very effective long term load estimation method. However the studies did not consider the same range of methods.

For this study we have applied a comparison of all known load estimation methods with the comparison focused on bias and error in order to test the robustness of loads methods as used by Kronvang and Brun, (1996), Littlewood, (1995), Preston, (1989). Secondly this study also focused on applying load estimation methods to runoff and constituent data sets of short duration and high frequency sampling. The performance of each load estimation technique has also been assessed against a range of sampling strategies.

2. METHODS

2.1. Load estimation techniques

The proliferation of load estimation techniques is partly due to the patchy nature of concentration data collection, the inconsistent correlation between constituent concentration and flow and between streams. For this paper we compare 34 load estimation methods from each of the three main load estimation categories;

- Averaging techniques (10 methods): some form of averaging is used in the concentration or flow data in the calculation of a load.
- Ratio method (14 methods): Load is determined based on the ratio of flow and concentration and often modified by a bias correction factor.
- Regression method (10 methods): based on fitting a relationship between flow and concentration for estimating a continuous trace of concentration.

For a full description of the 34 methods refer: Walling and Webb 1981, Littlewood, 1995, Letcher *et al*., 1999, Preston *et al*., 1989, Kronvang and Bruhn, 1996, Beale 1962, Quenouille, 1956, Tin, 1965, Rao, 1969, Ferguson, 1987.

2.2. Data Collation

The approach adopted for this study is similar to the methods used by Preston *et al* (1989), and Richards and Holloway (1987) who assessed long term data to determine annual mass loads. Similarly for this study a Monte Carlo analysis was used to determine the error in load estimation. The underlying timestep for the analysis is not annual, but on an event basis (duration 1-50 days) with average samples per event 15 ranging from 10-29 samples per event.

For each event, the existing concentration sampling regimes was assumed to perfectly represent the underlying concentration data. Each event was divided into 1500 equal duration time-steps and linearly infilled to have an equivalent concentration value for every time-step. We accept that this representation of the true concentration is only an approximation (but a limitation we cannot improve on with the available data) and that the base concentration data is also subject to uncertainty due to sampling and analysis methods.

High frequency flow data (hourly or better) was available for every event. The direct integration method was applied to the entire set of infilled concentration and flow data to determine the true load (as used by Preston, *et al.*, (1989)). This load estimate was treated as the 'true load' for comparison with load estimations determined with different combinations of concentration sampling routines and load estimation techniques.

In total, 31 events were identified as suitable. Sediment and nutrient data were provided for the majority of events, sediment data is only reviewed in this paper. Data sets were collected across Queensland (Figure 1) by those referred to in acknowledgements between 1976 – 2008. The minimum criteria for the data was that (a) at least 10 samples were collected across the runoff event (b) at least 3 samples had been collected on or before the peak of the runoff event (c) data was collected at an NRW gauging station where high frequency flow data was available. The concentration data was either manually sampled or using a pump sampler.

Figure 1: Queensland map indicating the spatial distribution of the 31 data sets used in the load estimation analysis.

Based on the long term ambient data for that site we set the 'before event' and 'after event' concentration values to reflect the long term ambient concentration values, this procedure is required to 'tie down' load estimation methods to correspond to the start and end of the runoff event.

2.3. Re-sampling

We re-sampled each of the 31 'synthetic' events comprising 1500 concentration values by randomly selecting concentration values from before or after the discharge peak. There were three distinct sampling strategies; sample the rise only, sample the fall only or sample on both the rise and fall. We took from 3 to 15 samples on either only the rise or the fall, and 3 to 30 samples that covered both the rise and fall; considering every combination of possible distribution of the number of samples on the rise and fall. The total number of subsample combinations was 250. Each sampling method was conducted 50 times, and for each of these (250x 50) for each event, 34 load estimation methods were applied before subsequent analysis of performance.

2.4. Statistical Methods applied to estimate error and confidence interval

In order to test the performance of the combinations of load estimation method, concentration sampling method and events, we have used the Root Mean Squared Error (RMSE) after Kronvang and Bruhn (1996), and Preston *et al*., (1989). The RMSE is made up of a measure of the variance and the bias of the estimated load. An ideal estimator should be both accurate (low bias) and be consistent in its prediction (low variance), hence our adopted measure of a robust result is a low RMSE value.

In addition to using RMSE to determine the most robust methods, we determined the 95% confidence interval when combining a given sampling method and load estimation method. This confidence interval was determined as 1.96 times the standard deviation of the error term, which is the mean deviation of the calculated load value (from 50 subsamples) from the true value for each event. The confidence interval is represented as a percentage of the true load.

3. RESULTS AND DISCUSSION

3.1. The most robust load estimation methods

Preliminary results demonstrate why there are so many alternative load estimation techniques; there is no clear best estimation technique across all sampling methods and events as illustrated by the varying percentages of most successful load estimation method for each of the sampling methods shown in Figure 2.

Where samples were collected on both the rise and fall, the ratio methods clearly outperformed the regression and averaging techniques where the sample size is up to 20 samples (Figure 2). Of the ratio techniques, Tin's modified ratio (method 19) (Tin, 1965), performed the best with the lowest RMSE 13% of the time for small sample sizes (3-7 samples), 21% for 8-12 samples, 26% for (13-20 samples and 24% for 21-30 samples, or an overall performance of being the lowest RMSE for 23% of the 224 sampling methods which included samples on both the rise and fall.

For sampling methods with samples collected on both the rise and fall of the hydrograph and sample number were greater than 20, then the averaging techniques performed the best. This is a logical result, as the averaging techniques essentially infill gaps in the existing concentration data (e.g. linear infill), and when the sample number is high and well distributed across the event then assumptions about the infilling method are less important because the event is well described by the sampled data.

The regression techniques performed poorly across all sample sizes when the samples were spread across the rise and fall. Regression techniques depend on a good correlation between flow and concentration, where this exists, the method performs well, and it appears from our analysis that around 20% of the events had a good correlation between flow and concentration. We have not analysed this result further to determine if the performance of the regression method can be simply related to physical variables such as geographic location, catchment area or dominant soil type.

Where samples are only collected on the rise, the averaging techniques performed very well with the lowest RMSE for approximately 50% of the analysis. If this is compared with the relatively poor performance of the averaging techniques where all samples are on the fall, we can conclude that the concentration samples collected on the fall of the hydrograph provide less information or are less valuable in determining the true load. The relative improvement in performance of the regression techniques where samples are taken only from the falling stage of the hydrograph, supports our observations that the correlation between flow and concentration has a better fit on the fall of a flood than on the rise. This is because the concentration peak usually occurs before the flow peak, hence there is a continuous decline in concentration values across the entire falling limb.

Figure 2 provides a useful summary for choosing the most robust load estimation method type to retrospectively apply to a given number of event samples, however it does not demonstrate the relative performance of each sampling strategy to help in the design of monitoring programs. Nor does it quantify the likely error of given sample method – load estimation method combinations. To do this we have further analysed the results to generate a lookup table to show the error and 95% confidence intervals for load predictions (Table 1). Table 1 only shows sampling strategies with up to seven samples to reflect the small sample size commonly observed in water quality data bases. Because we used a large number of events spanning a large geographic area, a large range in catchment areas and a range of shapes of hydrograph, the error and confidence interval values generated in Table 1 may be used as an indicative guide to the likely error when analysing new data sets which conform to the broad sampling categories used here. Table 1 can also be used as a guide to determine the most appropriate load estimation methods (smallest mean error and narrowest confidence interval in the error term which is represented by dark shading) for a given sampling technique. With this in mind we have excluded from Table 1 the results of sampling method – load estimation technique combinations where the method did not successfully return a load estimate within 1000% of the true load at least 80% of the time. We also excluded results which failed a power test of the number of Monte Carlo samples, where the mean of results from the fifty subsamples could not be determined to be within 20% the true subsample population mean (at the 0.05 level). This criteria is responsible for the large number of blank cells for low sample sizes. Most regression methods are entirely omitted because they did not demonstrate a suitable correlation to meet the above criteria (i.e. at least 80% of cases).

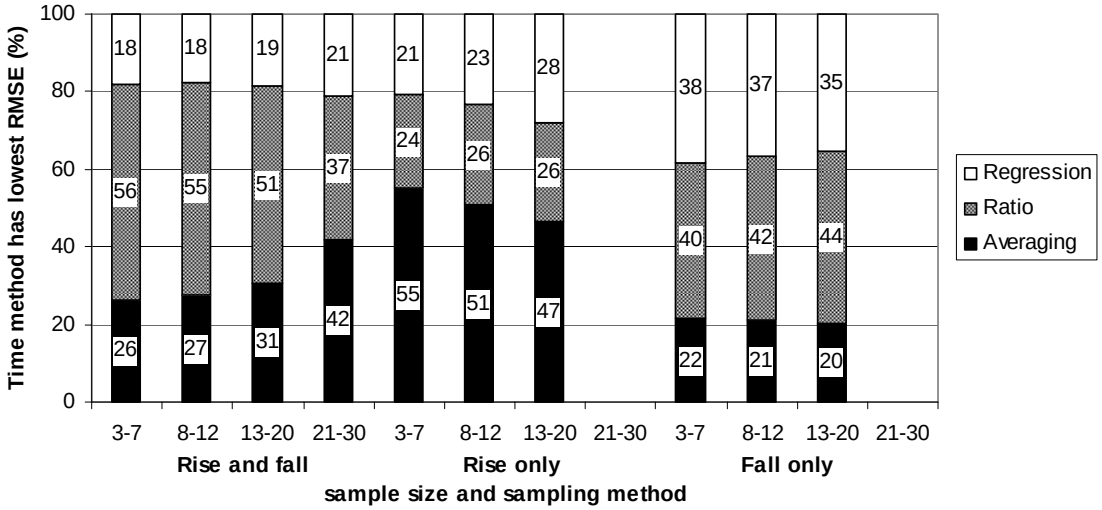

Figure 2: The Ratio methods produced the lowest RMSE when samples were collected on both the rise and fall and where the total number of samples was less than 21.

An important conclusion from Table 1 is that load estimation for sample sizes less than 8 rarely predict a load within 50% of the true mean (at 95% confidence level), and that the best performing sampling methods are those with samples on both the rise and fall of the event.

For specific streams where a tight correlation is known to exist one would expect regression methods to outperform the other methods, however with no prior knowledge of such correlations, then the advice gained from applying Table 1 will provide a conservative estimate of the error and 95% confidence interval in the calculated load.

4. CONCLUSIONS

Not surprisingly there is no clear best estimation technique across all sampling methods and events. The regression methods are robust where a tight correlation exists between flow and concentration, and such a relationship would tend to hold regardless of when and the frequency of sampling across the hydrograph. Unfortunately a minority of the events that we analysed had good flow and concentration relationship and as such this analysis indicates that regression methods would only be used in preference to other methods after a preliminary analysis has shown a good relationship between flow and concentration. Where there were at least some samples on the rise and fall the ratio methods were consistently more robust, being the best method in 56% of the cases for 3-7 samples If we use these results to consider the likely error in estimated load, the best sampling methods (considering only 3-7 samples) and load estimation method combinations can give a load estimate within 50% of the true load (95% confidence), although most combinations will fail to provide a load estimate within 100% or an order of magnitude of the true load.

Table 1: Mean error term (% of true load) ± 95% confidence interval (% of true load) of load estimation technique for sampling strategies of 3-7 samples, sample method combinations (only those methods that gave both a mean error and a 95% confidence interval <= 100% of the true load are shown). Shading indicated the combined mean error of our estimates plus the 95%confidence interval to give an estimate of the upper error range: ■ *<50% of the true load ,* ■ *50-75% of the true load* ■ *75-100% of the true load*

Sampling			**Averaging methods**								**Ratio methods**										**Regression**	
Rise	Fall	n	1	2	4	5	6	7	8	10	11	14	15	16	19	20	21	22	23	24	25	33
0	3	3		44 ± 31	38 ± 24						38 ± 24	37 ± 24		40 ± 24	38 ± 26							
0	4	4		44 ± 32	36 ± 24						36 ± 24	35 ± 24			36 ± 25							
0	5	5		43 ± 34	34 ± 24						34 ± 24	32 ± 23		35 ± 23	34 ± 26				34 ± 23			
0	6	6		43 ± 35	33 ± 26						32 ± 26	31 ± 24		34 ± 24	33 ± 28		33 ± 25		32 ± 23		43 ± 47	41 ± 44
0	7	7		44 ± 38	33 ± 29						33 ± 29	32 ± 29		35 ± 27	33 ± 30		33 ± 24		33 ± 28		43 ± 65	42 ± 63
3	0	3		37 ± 70																		
4	0	4		36 ± 74																		
5	0	5		35 ± 69																		
6	0	6		34 ± 65																		
7	0	7		44 ± 69																		
2	1	3		39 ± 47	43 ± 80						43 ± 80	46 ± 93			43 ± 80							
3	1	4		37 ± 46	44 ± 85						45 ± 85	47 ± 97		49 ± 99	44 ± 85							
4	1	5		39 ± 56	46 ± 98						46 ± 98				46 ± 98							
5	1	6		38 ± 54	47 ± 96					51 ± 91	48 ± 96				48 ± 96							
6	1	7		37 ± 54	47 ± 92					50 ± 98	47 ± 93	48 ± 99			47 ± 93						43 ± 93	45 ± 97
1	2	3		37 ± 30														35 ± 21		34 ± 22		
2	2	4		34 ± 36	37 ± 67	58 ± 67	58 ± 67	63 ± 80			37 ± 67	39 ± 79			36 ± 67			32 ± 24	42 ± 87	30 ± 23		
3	2	5		33 ± 40	38 ± 70	53 ± 55	53 ± 55	57 ± 68			38 ± 70	40 ± 77		39 ± 74	38 ± 70		47 ± 91	30 ± 23	42 ± 81	28 ± 22		
4	2	6		33 ± 45	40 ± 82	55 ± 77	55 ± 77	60 ± 95		45 ± 93	40 ± 82	41 ± 92		42 ± 90	40 ± 82			29 ± 24	43 ± 95	26 ± 22		
5	2	7		34 ± 51	40 ± 83	53 ± 74	53 ± 74	59 ± 90		45 ± 92	41 ± 83	42 ± 89		47 ± 100	40 ± 83			29 ± 23	43 ± 93	26 ± 21	36 ± 71	38 ± 75
1	3	4		37 ± 30	29 ± 21	45 ± 38	45 ± 38	45 ± 38			29 ± 22	30 ± 24		37 ± 33	29 ± 20			36 ± 28	32 ± 23	34 ± 28		
2	3	5		33 ± 28	32 ± 48	46 ± 41	46 ± 41	46 ± 46			32 ± 48	33 ± 57		34 ± 33	32 ± 48	27 ± 19	32 ± 28	31 ± 23	34 ± 59	28 ± 22		
3	3	6		31 ± 32	34 ± 59	42 ± 42	42 ± 42	46 ± 53		42 ± 70	34 ± 59	36 ± 67	35 ± 34	44 ± 77	33 ± 59	25 ± 18	40 ± 77	29 ± 24	37 ± 68	27 ± 22	37 ± 60	38 ± 61
4	3	7		32 ± 39	35 ± 69	44 ± 48	44 ± 48	48 ± 60	54 ± 85	41 ± 65	35 ± 69	37 ± 75	41 ± 99	38 ± 71	35 ± 69	22 ± 17	40 ± 85	28 ± 23	37 ± 77	25 ± 22	34 ± 79	34 ± 80
1	4	5		38 ± 30	28 ± 21	43 ± 40	43 ± 40	41 ± 39	45 ± 28		28 ± 21	28 ± 23			28 ± 22		31 ± 25	34 ± 24	28 ± 20	32 ± 24		
2	4	6		32 ± 26	30 ± 43	40 ± 39	40 ± 39	43 ± 46	44 ± 37		30 ± 43	31 ± 51		40 ± 62	30 ± 43	25 ± 18	35 ± 54	30 ± 22	32 ± 50	27 ± 21		
3	4	7	43 ± 32	31 ± 29	29 ± 49	36 ± 37	36 ± 37	39 ± 46	46 ± 47	37 ± 54	29 ± 50	31 ± 56	29 ± 29	38 ± 61	29 ± 50	22 ± 17	34 ± 60	28 ± 22	32 ± 56	25 ± 22	28 ± 29	28 ± 28
1	5	6		37 ± 29	28 ± 31	35 ± 33	35 ± 33	36 ± 35	43 ± 26		28 ± 31	29 ± 36		34 ± 44	28 ± 32		28 ± 20	33 ± 25	30 ± 36	31 ± 25		
2	5	7		34 ± 29	28 ± 40	33 ± 30	33 ± 31	35 ± 35	41 ± 32		28 ± 40	29 ± 46	28 ± 22	31 ± 29	29 ± 40	25 ± 21	32 ± 50	29 ± 24	30 ± 45	27 ± 23		
1	6	7		38 ± 34	26 ± 22	32 ± 28	32 ± 28	33 ± 30			26 ± 22	26 ± 23		29 ± 23	27 ± 24		28 ± 23	33 ± 26	27 ± 23	32 ± 27		30 ± 25

Table 2: Method number and name corresponding to Table 1

Num	Method	Num	Method	Num	Method
1	Flow x Concentration-	10	Flow stratified sampling	21	Goodman and Hartley Ratio-
2	Flow x Concentration all data-	11	Simple Ratio of average instantaneous flux and average flow-	22	Goodman and Hartley Ratio - flow stratified
4	Flow Weighted Concentration-	14	Beale Ratio-	23	Hartley Ratio-
5	Inter-sample mean concentration-	15	Beale Ratio - flow stratified-	24	Hartley Ratio - flow stratified-
6	Inter-sample mean concentration using Mean flow	16	Quenouille Ratio-	25	Concentration Power Curve Fitting-
7	Linear interpolation of concentration-	19	Tin's Modified Ratio-	33	Power Regression Curve Fitting Ferguson 1986
8	Average Between Sample Flow-	20	Tin's Modified Ratio - flow stratified-		

5. ACKNOWLEDGMENTS

The authors acknowledge the efforts of those who have contributed in some way to the collection of the water quality data used in this paper. These include staff from Natural Resources and Water, James Cook University, regional bodies and community groups across Queensland.

6. REFERENCES

Beale EML. 1962. Some uses of computers in operational research. *Industrielle Organisation* **31**: 51–52.

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

Ferguson, R.I (1987) Accuracy and Precision of Methods for Estimating River Loads, Earth Surface Processes and Landforms, Vol. 12, pp. 95-104.

Horowitz, A.J. (2003) An evaluation of sediment rating curves for estimating suspended sediment concentrations for subsequent flux calculations. Hydrological Processes, 17, pp. 3387-3409.

Keyes, A. M. and Radcliffe D. (2002) A Protocol for Establishing Sediment TMDLs. The Georgia Conservancy: Atlanta, GA; 31 pp.

Kronvang, B. and Bruhn, A.J. (1996) Choice of Sampling Strategy and Estimation Method For calculating nitrogen and phosphorus transport in small lowland streams, Hydrological Processes, Vol. 10, pp. 1483-1501.

Letcher, R.A., Jakeman, A.J., Merritt, W.S., McKee, L.J., Eyre, B.D. and Baginska, B. (1999). Review of Techniques to Estimate Catchment Exports, Environment Protection Authority, Sydney.

Littlewood, I.G. (1995) Hydrological Regimes, sampling strategies, and assessment of errors in mass load estimates for United Kingdom Rivers. Environmental International Vol. 21, No. 2 pp. 211-220.

Preston SE, Bierman VJ Jr, Silliman SE. 1989. An evaluation of methods for the estimation of tributary mass loads. *Water Resources*

Research **25**(10): 1379–1389.

Quenouille, M.H. (1956). Notes on bias in estimation. Biometrika 43, 353-60.

Quilbe, R., Rousseau, A.N., Duchemin, M., Poulin, A., Gangbazo, G., and Villeneuve. J. Selecting a calculation method to estimate sediment and nutrient loads in streams: Application to the Beaurivage River (Que´bec, Canada). Journal of Hydrology, 326, pp. 295-310.

Rao, J.N.K. (1969) Ratio and regression estimators. In North Carolina Symposium on the Foundations of Sample Survey Theory, ed. By N.L. Johnson and H.Smith, pp 213-34. New York; Wiley

Richards RP, Holloway J. 1987. Monte Carlo studies of sampling strategies for estimating tributary loads. *Water Resources Research* **23**(10): 1939–1948.

Tin, M. (1965) Comparison of some ration estimators, Journal of American Statistical Association, Vol. 60, No. 309, pp. 294-307

Walling, D.E. and Webb, B.W. (1981) The Reliability of suspended sediment load data, Erosion and Sediment Transport Measurement (Proceedings of the Florence Symposium, June) IAHS Publ. no 133

Webb, B.W., Phillips, J.M., Walling, D.E., Littlewood, I.G., Watts, C.D. and Leeks, G.J.L. (1997) Load estimation methodologies for British rivers and their relevance to the LOIS RACS® program. Sci. Total Environ. 194/195, pp. 379-389.

Young, T. C., DePinto, J. V., and Heidtke, T. M. (1988). 'Factors affecting the efficiency of some estimators of fluvial total phosphorus load', *Wat. Resour. Res.,* 24, 1535-1540.