Hydrological Sciences - Journal - des Sciences Hydrologiques, **33,** 5, 10/1988

Monitoring baseline suspended sediment in forested basins: the effects of sampling on suspended sediment rating curves*

ROBERT B. THOMAS

Pacific Southwest Forest and Range Experiment Station, Forest Service, US Department of Agriculture, 1700 Bayview Street, Arcata, California 95521, USA

Abstract Rating curves are widely used for directly assessing changes in the suspended sediment delivery process and indirectly for estimating total yields. Four sampling methods were simulated-over a 31-day record of suspended sediment from the North Fork of the Mad River near Korbel, California. The position and size of the four groups of plotted slope/intercept pairs indicated differences in bias and variance among the methods. Estimates of total yield for the 31-day period and for storms of three sizes were also biased according to sampling method. A standard bias-correcting technique improved yield estimates, but did not remove sampling bias uniformly. Methods of data collection have a large and systematic effect on the estimation of rating-curve parameters and on estimates of suspended sediment yield. Differences attributed to land management may, in fact, result from changes in sampling methods.

La surveillance des charges solides en suspension dans les basins forestiers avant traitement: les effecs de l'échantillonnage sur les courbes d'étalonnage des sédiments en suspension

Résumé Les courbes d'étalonnage sont largement utilisées, directement, pour mettre en évidence les changements dans le processus d'écoulement des charges solides en suspension, et aussi indirectment, pour estimer les débits totaux. Quatre méthodes d'échantillonnage ont été simulées, sur un relevé de 31 jours, des sédiments en suspension provenant de la branche nord du Mad River, prés de Korbel, en Californie. La position et la taille des quatre groupes de couples pente/segment de droite, en représentation graphique, ont révelé des différences de déviation systématique et de variance parmi les méthodes. Le calcul des débits totaux pour la période de 31 jours et pour des averses de hauteurs variées était biasé, lui aussi, selon la méthode d'échantillonnage. Une technique standard pour la correction de la déviation systématique a mieux

* Paper presented at the Workshop on Estimation of Natural Baseline Conditions as a Basis for Detecting Changes in Water Quality (HW8), Vancouver, British Columbia, Canada, August 1987.

précisé l'estimation des débits, mais elle n'a pas uniformément éliminé la déviation de léchantillonnage. Il se trouve que les méthodes de collecte des données ont un effet à la fois grand et systématique sur l'estimation des paramétres des courbes d'étalonnage, ainsi que sur celle des débits des charges solides en suspension. Les différences attribuées a l'aménagement des terres pourraient, en effet, découler des changements dans les méthodes d'échantillonnage.

INTRODUCTION

Background levels of suspended sediment in forest streams are routinely measured because changes in suspended sediment flux provide a simple means of monitoring forest management. The rate of sediment transport in forested streams can be affected by activities such as timber harvesting and road building. Also, other pollutants adhere to sediment particles, making suspended sediment a useful indicator of general water quality.

Levels of suspended sediment transport in streams must be determined before disturbance to assess change, but doing this properly is difficult. Background data sets and the estimates made from them are often suspect. Perhaps the most common reason is lack of adequate planning. The first need is to establish firm objectives for the study. The objectives must be specific enough to define analyses before the data are collected. Data collection is a process of quantification, and background (baseline) data are primarily a set of "control" measurements in a time-change experiment. Valid comparisons require some knowledge of the properties of the estimates made from the data. In particular, it is important to know the bias of the measurement and calculation procedures. Otherwise there is no way confidently to attribute detected differences to changed conditions. Suspended sediment rating "curves" (i.e. functions) are frequently used to define the water discharge/suspended sediment relationship. Changes can either be assessed directly by comparing "before" and "after" rating curves, or the curves can be used to estimate long-term or storm sediment yields. Several workers have discussed the problems of using rating curves (Walling, 1977a, b; Walling & Webb, 1981). These include precision and accuracy as well as the widespread use of diverse methods. Others have attempted to alleviate the problems (VanSickle & Beschta, 1983; Ferguson, 1986a, 1987) by accounting for sediment storage or by correcting transformation bias.

This paper reports a study of the effects of sampling methodology on the quality of rating curves and of their estimates of suspended sediment yield. In principle these ideas also apply to sampling other pollutant loads.

PROBLEMS IN COLLECTING SUSPENDED SEDIMENT DATA

Suspended sediment yield is estimated by measuring water discharge and suspended sediment concentration. The technology has long been available to measure discharge continuously, but suspended sediment concentration has traditionally been measured at discrete times.

Recent advances in instrumentation (Skinner *et al.*, 1986) allow automatic or continuous concentration measurement. These instruments generally require 120 V a.c. power, and can be expensive to purchase and operate. For small, remote, forested basins the method of choice is likely to remain collection of continuous discharge and discrete concentration data. Because the discharge data are essentially complete, the sampling problem reduces to deciding when to make the discrete measurements of concentration.

Most of the suspended sediment load in small forested basins is delivered during a very small portion of the time. The major flux of suspended material occurs during high flows when there is adequate stream power to transport the sediment (provided there is available sediment storage). This basic relationship is exploited in constructing rating curves.

A 31-day period in February and March of 1983 in the North Fork of the 10 442 ha Mad River basin near Korbel, in northern California, illustrates this point (Fig. 1). Ninety per cent of the suspended sediment was delivered in about 20% of the time. During an entire water year, it is not unusual for 95% of the suspended sediment to be discharged in about 5% of the time. This fact underscores the need to sample when the sediment flux is high. Extensive data sets with large portions of low-discharge sediment concentration data can contain less useful information than more limited sets collected during high flows.

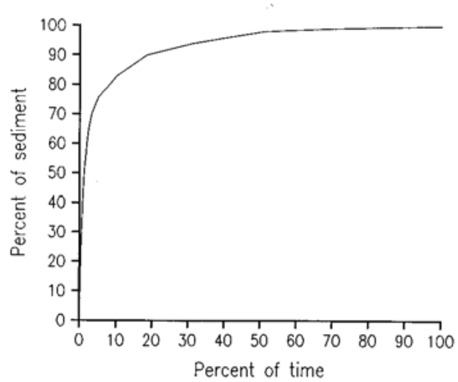

Fig. 1 *Percent of a 31-day period delivering percentages of suspended sediment yield. The data are from WY 1983 at the North Fork of the Mad River near Korbel, California.*

Suspended sediment data are often collected at fixed intervals (i.e. a systematic sample). This practice is appropriate (even required) if standard time series analyses are to be performed. The problem is one of balancing the need to sample adequately storm periods with the cost of frequent measurement. Usually, the sampling frequency required to characterize the important flows is much too high to be affordable during low flows. The frequency can be changed manually or otherwise to match the flows, but this complicates both the sampling operation and the estimation procedure. Also, it means that time series analyses based on equal-interval data cannot be applied.

Most studies of suspended sediment yields compare separate data sets. Background data may, for example, be compared with data collected after some land management activity or "treatment". These data are collected and summarized by a variety of methods. Just how these different sampling and calculation techniques affect the estimates is poorly understood.

SAMPLING FOR SUSPENDED SEDIMENT RATING CURVES

The data

The discharge hydrograph for the 31-day period shows storms of varying size (Fig. 2). The three storms marked are used in the following analysis. The Korbel station has line electrical power, and a continuously recording turbidimeter was installed. The turbidity/suspended sediment concentration relationship was determined empirically (R^2 = 0.99 for 52 observations, and the estimated standard error was 0.17% of the predicted mean concentration in mg l^{-1}). Data were selected from the turbidity chart at 4450 10-minute intervals and the concentrations estimated. The resulting sedigraph was obtained by multiplying each concentration by the corresponding water discharge and a constant to adjust units (Fig. 3). This plot again emphasizes the relative scarcity of times when measuring sediment flux is important.

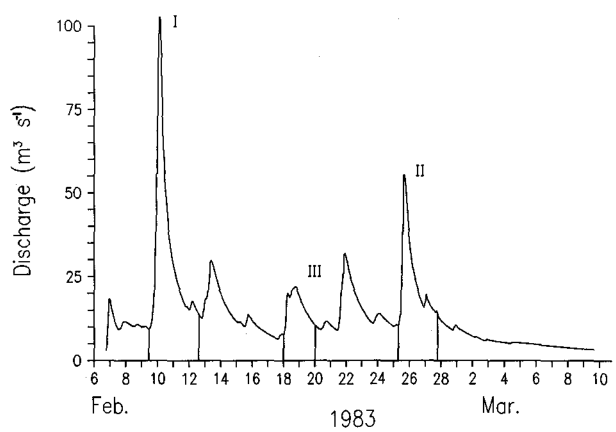

Fig. 2 *Hydrograph of a 31-day period from WY 1983 at the North Fork of the Mad River near Korbel, California. The three subperiods were selected to represent large, medium, and small storms for an analysis of the effect of storm size on estimation bias.*

Given the high data frequency, the suspended sediment flux is assumed to be known for the 31-day period. The 4450 points are treated as a

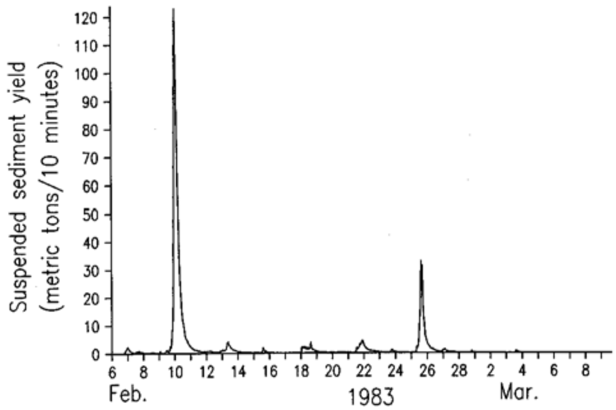

Fig. 3 *Sedigraph of a 31-day period from WY 1983 at the North Fork of the Mad River near Korbel, California. Suspended sediment yields were calculated for each 10-minute period.*

population from which several types of samples will be taken. The performance of different sampling schemes can then be compared to known "true" values obtained by summing the sedigraph for appropriate periods.

In addition to the hydrograph and sedigraph of the 4450 points, a scatter plot of the corresponding water discharge/suspended sediment concentration pairs was constructed (Fig. 4). Most of the data were from low

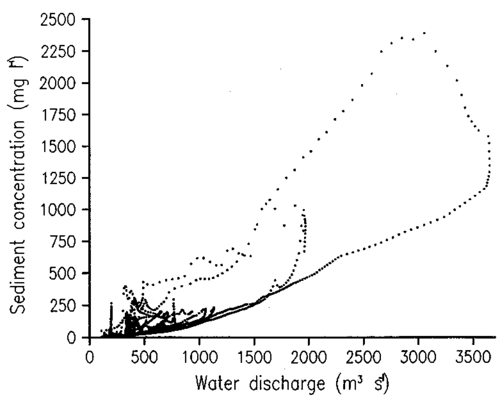

Fig. 4 *Simultaneously collected suspended sediment concentration/ water discharge pairs from the 31-day period at the North Fork of the Mad River near Korbel, California. There are 4450 points plotted. Hysteresis loops from the rises in discharge are evident.*

flows and show "hysteresis loops" produced by storms. The vertical "stacks" of points indicate rises in suspended sediment concentration during nearly constant flows. Sudden sediment contribution to the river at constant stage, such as by bank sloughing, is one possible explanation.

Data analysis

The suspended sediment rating data were analysed by taking natural logarithms of concentration and discharge (Fig. 5). (The scales were changed to powers of 10 for ease in reading.) This transformation has several effects. The points are more evenly distributed along the (log) discharge axis, which blends the effects of the hysteresis loops from varying size storms. The "swarm" of points is more linear although still concave upward. The vertical stacks of points at low discharges are relatively more prominent.

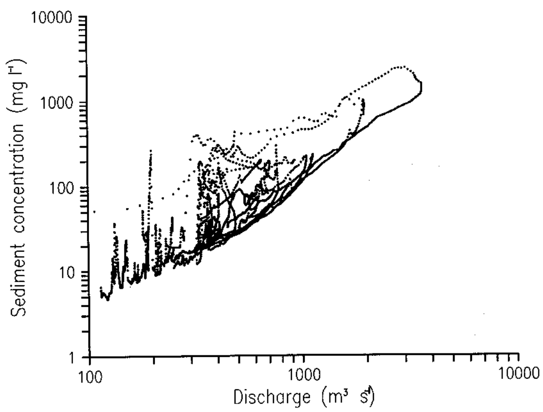

Fig. 5 *Natural log transformations of suspended sediment concentration and water discharge pairs shown in Fig. 4. The scales are shown in powers of 10 for ease in reading.*

Rating curves are usually formed by regression using least squares; however the form of the population shown in Fig. 5 does not encourage collecting samples to obtain least-squares estimates. The swarm of points is not linear, the vertical variation of points appears greater at the left than at the right, and the variation around any conceivable linear model is not normal. However, the very completeness of the data set may account for this pattern.

Most sets of rating data are based on less fequent observations over a longer time period. Separate storm features are less well defined, therefore, and data are included from more storms having different point patterns. The effect is to blur the individual storm patterns by selecting fewer points from each of them, and to heighten the illusion of normal random variation around a linear model by combining points from more differently-shaped events. The resulting pattern erroneously encourages using the regression approach.

In accordance with common practice, this population was sampled with four different sampling schemes to investigate their effects on estimates of sediment rating curves and yield.

Simulating four sampling methods

For convenience in nomenclature, standard statistical terminology was followed. The 4450 10-minute periods in the population are referred to as "sampling units". The aggregate of sampling units chosen by any of the sampling methods for analysis is called a "sample".

The four sampling methods can be divided into two classes. The first class is time-based, and consists of simple random and fixed-interval (or systematic) sampling. Each simple random sample was collected by making equal probability selections of 50 sampling units, without replacement. The systematic samples were collected by choosing one of the first 89 (4450/50 = 89) sampling units at random and then including every 89th one thereafter. This insured that 50 sampling units were in each systematic sample.

The second class of sampling methods is discharge-based. The first of these is flow-proportional. This method can be used when a station has instrumentation to estimate discharge in real time and can accumulate the total discharge since the last concentration sample. Samples are then taken at equal preselected intervals of accumulated discharge. This method intensifies data collection during high flows.

To simulate flow-proportional sampling, total water yield for the period was determined. The total yield was divided into 50 parts and a random point located in the first one. The concentration in the 10-minute period when this "random amount" of water had accumulated was used as the first sample unit. The remaining data were taken from the periods corresponding to increments of 1/50th of the total water yield from this starting point.

The second discharge-based method was Selection At List Time, or, SALT sampling (Thomas, 1985). SALT is a variable probability sampling scheme that enhances the probability of sampling high flows. The probabilities depend on estimates of yield from a rating curve or an average sampling-rate function. The SALT estimators generate unbiased estimates of total yield and variance. SALT is used here solely to collect rating-curve data; yield estimation is a direct application of the rating curve to the discharge record.

To operate SALT, a suspended sediment rating curve was determined for 100 points collected from an ***earlier*** record. This procedure simulates information that would be available at the time of sampling. SALT sampling is governed by a set of random numbers preselected and stored in a computer at the gauging station. A new set of random numbers was created for each of the SALT samples in these simulations. Sampling parameters were set to obtain 50 sampling units, but, because SALT sample sizes are random, not all samples were exactly of size 50.The mean size for the 25 simulated SALT samples was 49.5 with a standard deviation of 0.7.

Twenty-five samples of size 50 (approximately 50, for SALT) were created for each of the four sampling methods. Natural logarithmic transformations were calculated for both the selected concentrations and corresponding water discharges, and rating-curve regressions were performed for each set. The least-squares estimated slopes were then plotted against the corresponding intercepts (Fig. 6).

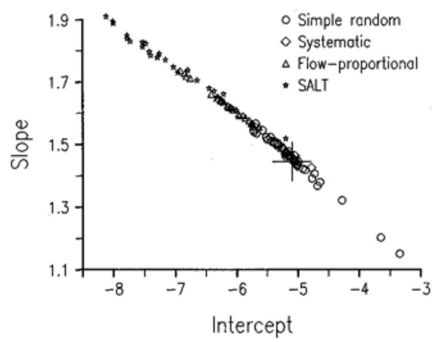

Fig. 6 *Slope/intercept pairs from regressions calculated from simulated samples selected from the 4450 point population according to four sampling methods. The "plus" symbol indicates the slope/intercept pair from the regression of all 4450 points.*

RESULTS

Rating curves

The swarms of rating-curve slope/intercept pairs indicate the performance of each sampling method. The time-based sampling methods tend to have larger intercepts and lower slopes than the discharge or sediment-based methods (Table 1). They also have nearly identical means of slopes and intercepts. The simple random sampling pattern exhibits larger variation than systematic sampling. The intercept and slope from the "rating regression" of all 4450 points in the population (indicated by the "plus" sign) are almost identical with the corresponding means of the "time-based" sampling methods (Table 1). The flow-proportional regressions tend to have smaller intercepts and larger slopes, especially those for the SALT simulations.

Table 1 *Means and standard deviations (SD) for rating curve slopes and intercepts from simulations of four sampling methods and the population rating regression*

	Slope:		*Intercept:*	
Sampling method	*Mean*	*SD*	*Mean*	*SD*
Simple random	*1.45*	*0.106*	*-5.07*	*0.623*
Systematic	*1.47*	*0.020*	*-5.13*	*0.136*
Flow-proportional	*1.63*	*0.052*	*-6.17*	*0.362*
SALT	*1.76*	*0.099*	*-7.07*	*0.739*
Population regression	*1.46*	*X*	*-5.10*	*X*

The points from all of the methods as well as the separate simulations lie nearly on a straight line. Part of this pattern can be understood from the rating data plot (Fig. 5). Most of the data lie near the bottom of the point swarm which is concave upward. The time-based methods preferentially select points towards the left of the range of discharge which tends to reduce the slope and increase the intercept. The discharge-based methods tend to select points to the right, reversing this pattern.

The slope/intercept pairs lie near a straight line due to a mathematical property. Consider a family of straight lines passing through a single point on the rating data plot. The lines have different slopes and intercepts, but all have the one point in common. The slopes can be expressed as a linear function of the intercepts with parameters that depend on the ***coordinates of the common point***. Each ***line*** through the point on the rating data plot (Fig. 5) can therefore be expressed as a ***point*** on the slope/intercept graph (Fig. 6), all of which lie on the same straight line.

The rating curves do not pass through a single point on the rating data plot, but most data sets will tend to have similar means of (log) discharge and (log) concentration. Since all simple linear regressions go through the coordinate of the sample means, the rating curves will go through a small region near the coordinate of the population means. Therefore the slope/intercept coordinates will tend to be located on or near a straight line.

Yield estimates for the four sampling methods

Rating curves are often applied to water discharge records to estimate total suspended sediment yield for a period. The 25 sampling rating curves for each of the four methods were used to estimate suspended sediment concentration for each of the 4450 points. The estimated concentration for each 10-minute period was then multiplied by the corresponding water discharge and a unit-adjusting constant to estimate the 10-minute sediment discharge. These estimated yields were summed and plotted (Fig. 7). Each short vertical line represents one simulation, and the longer vertical lines show the means of the 25 simulations for that method. The "true" total suspended sediment discharge for the entire period is 8306 t and is indicated by the longest vertical line traversing all four methods. The difference between the mean for each sampling method and the "true" discharge will be used to estimate the bias.

The simple random sampling simulations underestimate the "true" yield between 18 and 60% with a mean negative bias of about 35%. Systematic sampling gave slightly higher values with a mean negative bias of about 31%, but with greatly reduced variation. Discharge-based methods gave means closer to the "true" yield; the flow-proportional data are biased about 19% low and the SALT data around 13%. The flow-proportional data exhibit considerably less variation than the SALT data.

None of these methods is satisfactory. It appears that estimators of suspended sediment yield are biased using any of these sampling methods and that the bias depends (as does the standard deviation) on which method is used. This assessment is consistent with other work investigating bias in estimating suspended sediment yield using sediment rating curve methods (Walling, 1977a, b; Walling & Webb, 1981).

The "rating curve" from all 4450 points in the population has slope and intercept close to those for the time-based sampling methods. This implies that increasing the sample size for those methods would not reduce the bias. It is disconcerting when an estimating procedure does not produce the correct value even when applied to the entire population.

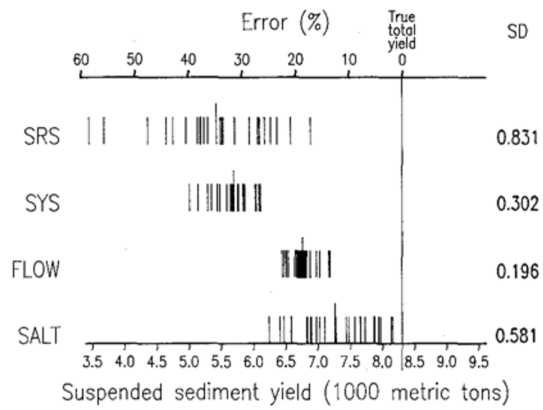

Fig. 7 *Estimates of suspended sediment yield and their standard deviations (SD) from the 31-day period for four different sampling methods. Each short vertical line represents one simulated sample. Means for the estimates of each method are indicated by the slightly longer lines. The "true" total yield is indicated by the long vertical line extending through all sampling methods. These estimates are not bias corrected.*

One tacit assumption in analysing rating curve data is that the least-squares regression residuals in the ***transformed*** space are normally distributed. In this case, back-transforming to estimate concentration gives the median and not the mean. To estimate the mean it is necessary to use a bias correction technique. Suppose the overall variance estimated by the rating regression is denoted by s^2. Then $\exp(s^2/2)$ is a reasonable bias correction factor for the natural log transformation (Brownlee, 1967). Multiplying the usual back-transformation by this factor gives an approximate correction to the mean.

Several investigations have applied this bias correction technique to sediment rating curves (Thomas, 1985; Ferguson, 1986a, b, 1987; Koch & Smillie, 1986). The detailed data from Korbel illustrate how this bias correction procedure behaves for the four sampling methods (Fig. 8).

The estimated bias is reduced for all methods, ranging from about 23% low for the simple random sampling method to only 2% for the SALT method. However, all of the standard deviations are increased. This will generally be the case. In a lognormal distribution the median is less than the mean, s^2 is non-negative, and the bias correction is multiplicative. The

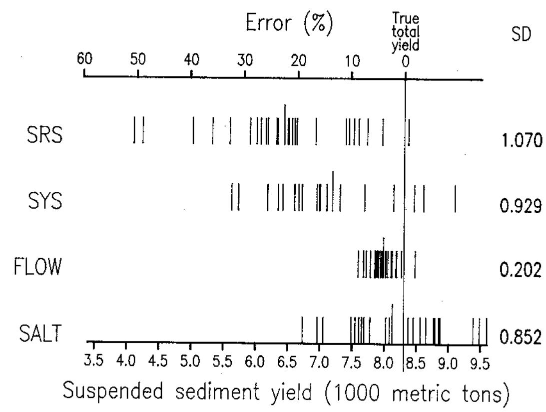

Fig. 8 *Bias corrected estimates for the data presented in Fig. 7. The standard correction for lognormal transformations was used.*

correction factors will therefore displace the estimates to the right (except when the estimate of variance is zero) and tend to spread them out, thereby increasing the variance. The spread of the corrected values will not always be greater than the spread of the uncorrected values, however, because the correction factors are different for each regression.

While the sets of corrected estimates are closer to the "true" yield, problems still remain. The variances of the estimates have been increased, although they could be reduced by increasing the sample sizes. More importantly, the means of the estimates for the four methods still do not coincide with the "true" value, although those for the flow proportional and SALT methods are close. That is, the bias is not removed in general by this procedure, and the bias before and after the correction still depends on the sampling method.

Estimating storm yields

To illustrate how the sediment rating curve method performs when estimating storm yields, three storms were selected from the period to represent large, medium, and small events (Fig. 2). The 50-observation systematic and flow-proportional sampling simulations were used to estimate suspended sediment yields for the three storms. The same 25 simulations developed for the rating curve and total yield calculations were used. The estimates were again obtained by applying the appropriate sediment rating curves to the 10-minute discharges in the three storms. The "true" storm yields and the means of the simulations for the two sampling methods, both uncorrected and bias corrected, were compared (Table 2).

The means of the uncorrected estimates for both simulated sampling methods are less than the "true" values for all storms. The estimated negative bias for the systematic sampling method is 38, 34, and 19% for storms I, II, and III respectively. The comparable values for the flow-proportional method are 21, 24 and 18%. Evidently, the percent underestimation is greater for larger storms, especially for the systematic

Table 2 *"True" yields and uncorrected (U) and bias-corrected (BC) means of suspended sediment yield estimates of three storms using the systematic and flow proportional sampling methods (kt)*

	Systematic:		*Flow-proportional:*		*"True"*
Storm	*U*	*BC*	*U*	*BC*	
1	*3.43*	*4.33*	*4.34*	*5.14*	*5.51*
11	*0.928*	*1.17*	*1.06*	*1.26*	*1.40*
III	*0.211*	*0.266*	*0.214*	*0.254*	*0.262*

sampling method. This is consistent with the underestimated yields for the entire 31-day period.

While this general relationship is suggested for these storms, it could be confused by two factors. One is that the distribution of flow duration for storms of a given size (i.e. volume or peak) can be quite different. The interaction of the rating curve estimation procedure with different distributions can conceal the relationship. Another factor is that of sediment storage in the channel due to the timing of storms relative to preceding ones of varying sizes. For a storm immediately following a large storm the channel may be "starved" of sediment and the rating curve would tend to overpredict the actual sediment load. The opposite might hold for the first storm after a significant dry period, or a storm closely following other small ones.

The bias-corrected estimates do somewhat better. Bias for the systematic sampling method drops to a negative 21 and 16% for storms I and II respectively, while that for the flow-proportional method drops to a negative 7 and 10%. The bias of both sampling methods for storm III essentially disappears. While encouraging, it is not clear what properties of storm III produced this result. Other storms of the same "size" could have different levels of bias correction due to the factors mentioned above. Again, the bias correction procedure does not perform properly for all situations.

Estimated standard deviations for uncorrected and bias-corrected estimates were calculated (Table 3). The standard deviations increase for every case except the flow-proportional method in storm I, where it drops slightly. The rise in standard deviation is much greater in general for the systematic sampling method than for the flow-proportional method.

Table 3 *Estimated standard deviations for uncorrected (U) and bias-corrected (BC) estimates of suspended sediment yield for three storms and the systematic and flow proportional sampling methods (kt)*

	Systematic:		*Flow proportional:*	
Storm	*U*	*BC*	*U*	*BC*
1	*0.195*	*0.562*	*0.207*	*0.178*
II	*0.048*	*0.152*	*0.019*	*0.033*
III	*0.010*	*0.035*	*0.006*	*0.013*

DISCUSSION

These results have consequences for using rating curves directly for assessing changes in suspended sediment delivery regimes. Rating curve position (as indicated by the regression parameters) is evidently biased depending on the method of sampling as well as on physical characteristics of the catchment being measured. Therefore, it may not be clear that a detected change is due to treatment. This could explain the high variation seen in rating curve position under the "same" conditions in the same stream (Beschta 1981).

If rating curves must be used to assess change, it is essential to use the same sampling practices for both baseline and post-treatment measurements. This procedure may not be simple. The simulated sampling methods used in this study were idealized and applied rigorously. Field data collection is less carefully controlled, varying due to changes in instrumentation, budgets, and personnel. The effect of such interacting factors on the bias of the estimates is problematic.

Using a sampling method having low variance is another desirable feature when comparing rating curves. The flow proportional method has the lowest variance for these data as well as fairly low bias. The bias, however, is likely to depend on the particular situation, and may change after treatment. Also, the bias will not be generally known. This, combined with doubt about the meaning of the variance estimate for a poorly specified regression model, raises serious questions about any formal comparisons made.

Rating curves can also estimate suspended sediment yields using various techniques, depending on the discharge data available and the preferences of individual workers. Applying a rating curve directly to detailed discharge data as done in this paper should give the best results for the method. Also, these rating data were taken from the period for which yields were estimated, which is often not the case. The yields estimated with these data, however, did not perform well. It is to be expected that the yield estimates depend on sampling method as did the rating curves themselves. In addition, however, the bias in the yield estimates may depend on the level of flows being estimated. Using the same sampling method to equalize bias may not ensure valid comparisons. Differences in storm patterns, over which the investigator has no control, could continue to influence the results.

The discharge-based sampling methods may perform somewhat better when simulated than under field conditions. In a field setting the discharge pattern is not known before measurement. Sampling parameters cannot, therefore, be set to collect a preset sample size. This may increase the variance of the estimates. This source of variability is not present in the simulated time-based sampling estimates.

Probability sampling is recommended for estimating suspended sediment yield. With probability methods the relationship between sampling and estimation is understood. If the methods are applied correctly the properties of the estimates are known (e.g. unbiasedness). Also, every sample allows an estimate of variance. These properties are all known from theory and can be applied generally without requiring extensive and repeated empirical testing.

Probability methods are sometimes considered too difficult (i.e. too costly) to apply under most hydrological field conditions, largely due to stringent requirements for random sampling. It is reasonable to ask, however, if such difficulty is any greater than that resulting from methods of sampling that are easier to apply, but whose properties are not understood. Estimates having uncertain properties can cast doubt on the study results themselves. The benefits of a sampling method should be considered as well as its costs.

The standard bias correction procedure for lognormal transformation did not improve the estimates uniformly. The improvement clearly depended on the sampling method and intensity as well as on the set of discharges in the period being estimated. The performance of this correction procedure is sensitive to the assumptions, particularly to that of normality in the log space (Duan, 1983). The assumption of normality is hard to justify considering how basic rating data are produced, and poor performance is to be expected. Therefore, blanket application of this bias correction procedure cannot be recommended.

While these results may be unsettling, especially due to the widespread use of rating curves, they should not be surprising. Sampling methodology is known to be important in making estimates that have acceptable properties and for which valid estimates of error can be calculated. The required relationship between sampling procedures for suspended sediment rating data and methods of yield estimation has never been demonstrated. This fact is evidenced by the large number of techniques used to sample data for rating curves and to calculate yields. The methods seem to have evolved more from convenience and intuitive guesswork than from rigorous analysis. Clearly, rating curve estimates are systematically tied to the way the data are collected and analysed.

The basic problem is that the model is not adequately specified. There is certainly a positive correlation between water discharge and simultaneously collected suspended sediment concentration in most streams. There is not enough information in the discharge, however, to predict concentration in the commonly used simple linear regression context.

These results apply strictly only to the set of data investigated. Many of the characteristics in this data set, however, are present in other sets as well. Hysteresis loops, the general pattern of the swarm of data points, and the shape of the transformed swarm are all features common to these sets of data. It is hardly to be expected that other rating data sets will have the properties required to make their estimates behave correctly.

CONCLUSIONS

Rating curves are widely used by hydrologists to assess changes in suspended sediment regimes and to estimate yields for specified periods. Tacit in this practice is the assumption that any measured differences from baseline conditions are due to changes in the underlying sediment production mechanism, possibly obscured by random variation. The results of simulated sampling applied to a 31-day record having essentially complete sediment information show that other factors affect the estimates:

- Four sampling techniques gave ***systematically*** different patterns of suspended sediment rating curve position (i.e. rating curve parameters).
- Estimates of total suspended sediment yield for the entire 31-day period and for selected storms were also ***systematically*** altered depending on the sampling method used.
- Storm estimates appeared to be dependent on storm size, with the percent underestimation generally increasing for larger storms. This effect, however, could be confounded by storm order and timing interacting with available channel sediment.
- The standard bias correction for lognormal transformations was applied to all of the yield estimates. The bias was reduced, but was still dependent on the sampling method employed and on storm size. The variance of the corrected estimates was generally increased.

Assuming that changes in rating curves come solely from changes in the process producing suspended sediment seems unwarranted. Sampling methods also appear to influence systematically the estimates of the rating curve in complex ways. There is also doubt about exactly what a rating curve represents. Rating curve predictions of suspended sediment discharge are evidently dependent on information not available in the explanatory variable. Hydrologists should work to develop better models for predicting concentration, models that include sampling as an essential component.

Acknowledgements I thank Diane L. Rainsford for her exceptional efforts in writing the computer programs for this study and for preparing the figures. I also thank the Simpson Timber Company for its cooperation with the Forest Service, US Department of Agriculture, in collecting data at the test site.

REFERENCES

Beschta, R. L., O'Leary, S. J., Edwards, R. E. & Knoop, K. D. (1981) Sediment and organic matter transport in Oregon Coast Range streams. *Rep. WRRI--70, Water Resour. Res. Inst., Oregon State Univ., Corvallis.*

Brownlee, K. A. (1967) *Statistical Theory and Methodology in Science and Engineering,* 2nd edition. John Wiley & Sons, New York.

Duan, N. (1983) Smearing estimate: a nonparametric transformation method. *J. Am. Statist. Ass.* **78** (383), 605-610.

Ferguson, R. I. (1986a) River loads underestimated by rating curves. *Wat. Resour. Res.* **22** (1),74-76.

Ferguson, R. I. (1986b) Reply to 'Comment on "River loads underestimated by rating curves" by R. I. Ferguson'. *Wat. Resour. Res.* **22** (13), 2123-2124.

Ferguson, R. I. (1987) Accuracy and precision of methods for estimating river loads. *Earth Surf. Processes and Landforms* **12** (1), 95-104.

Koch, R. W. & Smillie, G. M. (1986) Comment on "River loads underestimated by rating curves" by R. I. Ferguson. *Wat. Resour. Res*. **22** (13), 2121-2122.

Skinner, J. V., Beverage, J. P. & Goddard, G. L. (1986) Continuous measurement of suspended-sediment concentration. In: *Fourth Federal Interagency Sedimentation Conference* (Las Vegas, Nevada), vol. 1, 29-39.

Thomas, R. B. (1985) Estimating total suspended sediment yield with probability sampling. *Wat. Resour. Res.* **21** (9), 1381-1388.

VanSickle, J. & Beschta, R. L. (1983) Supply-based models of suspended sediment transport in streams. *Wat. Resour. Res.* **19** (3), 768-778.

Walling, D. E. (1977a) Assessing the accuracy of suspended sediment rating curves for a small basin. *Wat. Resour. Res.* **3** (13), 531-538.

Walling, D. E. (1977b) Limitations of the rating curve technique for estimating suspended sediment loads, with particular reference to British rivers. In: *Erosion and Solid Matter Transport in Inland Waters* (Proc. Paris Symp., July 1977), 34-48. IAHS Publ. no. 122.

Walling, D. E. & Webb, B. W. (1981) The reliability of suspended sediment load data. In: *Erosion and Sediment Transport* (Proc. Florence Symp., June 1981), 177-194. IAHS Publ. no. 133.

Received 23 January 1988; accepted 6 May 1988.