ELSEVIER

Journal of Hydrology 170 (1995) 27-45

Journal of Hydrology

[1]

# An evaluation of flow-stratified sampling for estimating suspended sediment loads

Robert B. Thomas, Jack Lewis*

*Pacific Southwest Research Station, Forest Service, US Department of Agriculture, 1700 Bayview Street, Arcata, CA 95521, USA*

Received 14 June 1994; revision accepted 31 January 1995

## Abstract

Flow-stratified sampling is a new method for sampling water quality constituents such as suspended sediment to estimate loads. As with selection-at-list-time (SALT) and time-stratified sampling, flow-stratified sampling is a statistical method requiring random sampling, and yielding unbiased estimates of load and variance. It can be used to estimate event yields or to estimate mean concentrations in flow classes for detecting change over time or differences from water quality standards. Flow-stratified sampling is described and its variance compared with those of SALT and time-stratified sampling. Time-stratified sampling generally gives the smallest variance of the three methods for estimating storm yields. Flow-stratified sampling of individual storms may fail to produce estimates in some short-lived strata because they may have sample sizes of zero. SALT will tend to give small samples and relatively high variances for small storms. For longer and more complex hydrographs, having numerous peaks, flow-stratified sampling gives the lowest variance, and the SALT variance is lower than that of time-stratified sampling unless the sample size is very large. A desirable feature of flow-stratified sampling is that the variance can be reduced after sampling by splitting strata, particularly high flow strata that have been visited just once, and recalculating the total and variance. SALT has the potential to produce the lowest variance, but cannot be expected to do so with an auxiliary variable based on stage.

## 1. Introduction

Standard references on suspended sediment methods (US Interagency Committee on Water Resources, Subcommittee on Sedimentation, 1963; Guy and Norman, 1970; Vanoni, 1975) generally emphasize measurement techniques for obtaining

---

* Corresponding author.

*SSDI* 0022-1694(95)02699-1

instantaneous concentrations, and offer minimal guidance for estimating event loads or seasonal loads. Rules for sampling-that is, selecting times to measure concentrations-are limited to generalities and the logical connection between sampling procedures and load estimates is largely ignored. Because measuring concentrations is costly, methods are needed which effectively use limited sets of concentration data. The selection and application of these data can have profound effects on the estimated suspended sediment yield.

A widely used method involves a sediment rating curve which expresses sediment concentration as a continuous function of discharge based on a sample of concentration-discharge pairs (Campbell and Bauder, 1940; Miller, 1951; Glysson, 1987). Predictions from a rating curve are applied either to flow duration data or directly to the hydrograph. These methods have a long history of development and use owing to their relative ease of application and flexibility.

Comparing rating curve estimates and more direct and intensive measures of suspended load show that rating curve methods can be biased and highly variable (Colby, 1956; Walling, 1977a,b). Also, rating curve estimates depend on sampling protocol (Bennett and Sabol, 1973; Beschta, 1978; Thomas, 1988). Methods for correcting bias in logarithmically transformed data (typically used for rating curves) have been suggested as a solution (Ferguson, 1986, 1987). However, although bias correction is beneficial, it does not rectify problems in estimates obtained from poorly fitting models (Thomas, 1985, 1988; Koch and Smillie, 1986; Walling and Webb, 1988). Cohn et al. (1992) stated that some bias-corrected estimates appear not to be seriously affected by lack of fit, but this was shown for a sample of sediment rating curve data rather than a complete hydrograph. Sediment concentration is clearly dependent on factors other than simultaneous discharge (Rieger and Olive, 1984), and rating curves rarely, if ever, reflect watershed dynamics adequately to give reliable yield estimates.

An alternative to using rating curves is to sample concentration much more frequently with automated pumping equipment or to estimate concentration from a highly correlated surrogate variable, such as turbidity, which can be measured continuously (Rieger and Olive, 1988). Instantaneous measurements or estimates of concentration are then multiplied by discharges representing short time periods and the products are summed over all time periods to obtain the load (Porterfield, 1972). Either of these methods may serve in particular applications, especially for research studies requiring detailed information on watershed response to treatments. However, the expense of collecting and analyzing much larger sets of concentration data and difficulties associated with remote data collection rule out this approach in other cases.

In rivers where concentrations change slowly over time another sampling strategy measures concentrations at longer intervals and uses time series analyses to estimate loads (Gurnell and Fenn, 1984; Lemke, 1991). However, time series models require equally spaced measurements and, in streams where data must be closely spaced to define occasional short periods of high sediment delivery, the cost can be prohibitive. Sampling strategies that vary sampling rates to reflect changing load conditions are needed to reduce cost without sacrificing accuracy and precision.

Although error estimates have not traditionally been reported for estimates of suspended sediment load, Gilroy et al. (1990) described methods for estimating the error of rating curve load estimates. They cautioned, however, that these methods require knowledge of the true model form and coefficients, and can be severely biased when the model is extrapolated beyond the range of the developmental data.

By defining a time-based population and sampling at random, unbiased estimates of both load and variance can be obtained with standard formulae from sampling theory. Stratified or variable probability sampling plans can reduce variance by taking advantage of population structure. A programmable data logger can direct random sampling by an automatic pumping sampler and adjust sampling intensity to real-time flow conditions. The additional effort needed for random sampling is repaid by estimates that, neglecting measurement errors, are unbiased regardless of the sample and population to which they are applied. Selecting a plan depends primarily on ease of use and magnitude of variance. Two such plans have been proposed recently (Thomas, 1985; Thomas and Lewis, 1993). We describe a third plan here and compare its performance with that of the other two.

## 2. Sampling designs

### 2.1. Suspended sediment populations

In suspended sediment sampling, the population of interest is typically the sediment load for a hydrograph or some partition thereof. The sampling units are defined as sediment loads for some short fixed interval of time, short enough to be adequately characterized by a midpoint discharge and concentration measurement. (We assume in this paper that stage and water discharge are available in real time and that sediment concentration is determined later by laboratory analysis. We also assume that the concentration data are accurate measures of, or can be transformed with minimal error to, cross-sectionally averaged concentrations. This must be verified in practice.) We have found 10 min intervals appropriate for small ‘flashy’ streams, but longer ones may be satisfactory for larger and more stable rivers. We use the term ‘unit yield’ for the mass of sediment discharged in one interval. Our objective is to specify the most efficient (i.e. minimal variance for a given sample size) sampling methods for estimating total sediment yield (sum of the unit yields) for various types of hydrographs.

### 2.2. Selection-at-list-time (SALT)

SALT is a variable probability sampling method similar to PPS (probability proportional to size) sampling (Hansen and Hurwitz, 1943). Their estimation formulae are identical. Both methods utilize an auxiliary variable, easily measurable for the entire population, to assign selection probabilities to each unit of the population. The variance is minimized for auxiliary variables that are proportional to the variable of interest. PPS requires enumerating the population and measuring the auxiliary variable on the whole population before sampling. SALT was developed as an alternative to PPS for populations which cannot be revisited for sampling (Norick, 1969). Estimating the SALT auxiliary variable total in advance permits sampling probabilities to be computed on the first pass through the population. Immediately upon measuring each unit’s auxiliary variable, a decision is made whether or not to select the sampling unit. The auxiliary variable might be a stage-based prediction of unit yield from a sediment rating curve (Thomas, 1985). This tends to concentrate measurements of sediment during high flows and continuously adjusts sampling intensity over a wide range of conditions. To restrict sample sizes, other functions of stage have also been used (Thomas, 1989).

### 2.3. *Time-stratifed sampling*

In time-stratified sampling for sediment yield, the hydrograph is partitioned into intervals of time (consisting of integral numbers of sampling units) called strata (Thomas and Lewis, 1993). Stratum lengths are determined at the start of each stratum by the current stage height and direction. An automatic pumping sampler is operated at a simple random sample of times from each stratum, and sediment yield is estimated by the usual stratified sampling formula for the total of a finite population (Eq. (1) below). By keeping strata short and allocating only two measurements to each stratum, time-stratified sampling has given estimates of load having lower variance than SALT, especially for individual storm events. Time-stratified sampling temporally distributes samples more evenly than SALT, insuring adequate sample sizes and hence good precision for small storms as well as large ones.

### 2.4. *Flow-stratified sampling*

Flow-stratified sampling stratifies a hydrograph by water discharge rather than time. The range of flow is divided into classes by stage height and direction and each flow class is randomly sampled during the time it is occupied. In contrast to time-stratified sampling, non-contiguous samples from the same flow class are placed in the same stratum. Because there are fewer strata, sample sizes can potentially be smaller, even when classes are divided by the same stages that govern time-stratum lengths.

Fig. 1 illustrates flow-stratified sampling with class boundaries at stages of 130 and 180 cm. In this example, the class definitions are identical for rising and falling stages, but they need not be. Vertical lines in the figure show the partitioning of the sedigraph at class boundaries. It should be noted that class boundaries include stage reversals. To detect a stage reversal, the hydrograph must drop a specified amount, $c_p$, below the preceding peak or rise an amount, $c_t$, above the preceding trough (Fig. 2). This procedure avoids frequent shifting between rising and falling strata when a hydrograph ‘wanders’ near peaks and troughs. Sampled units representing the same stratum in Fig. 1 are denoted by the same symbol. The eight partitions are combined into six strata, as two classes are visited twice.

Standard stratified sampling procedures must be altered, as the stratum sizes are

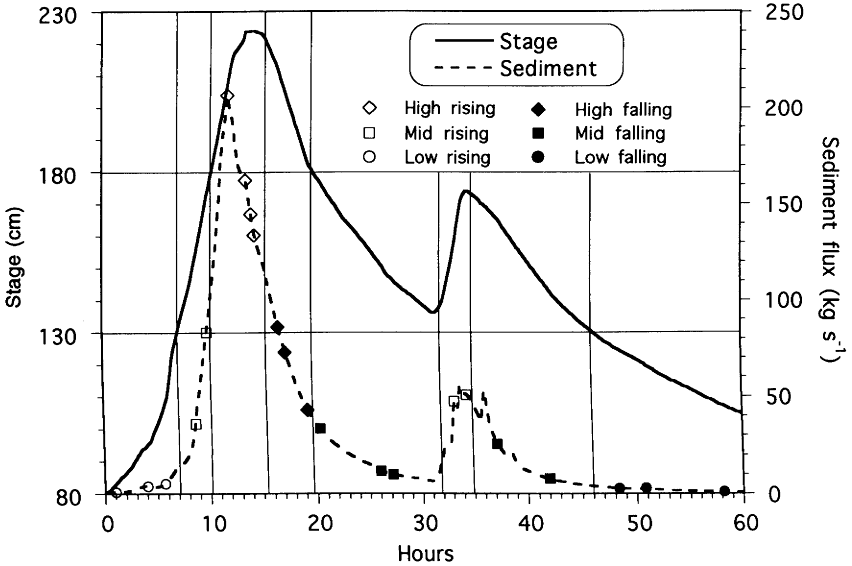

Fig. 1. Typical flow-stratified sample for three rising and three falling stage classes with lines indicating stage and sediment flux. Symbols identify stratum membership for each sampling unit selected for measurement. Rising and falling stratum boundaries need not be the same; they were made equal here to simplify the drawing.

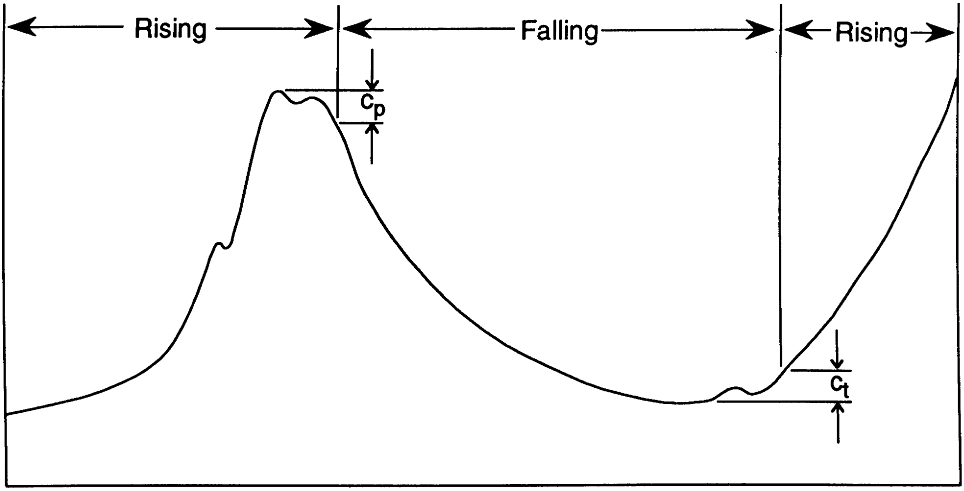

Fig. 2. Typical hydrograph to illustrate tolerances at peaks, $c_p$, and troughs, $c_t$, used to establish reversal of stage condition. For reversal detection, stage must drop more than $c_p$ below the maximum stage or rise more than $c_t$ above the minimum stage retained since the last reversal. Reversal detection follows shortly after peaks and troughs.

not known until the end of the monitored period. Sampling decisions must be made 'list-sequentially', i.e. as each population unit materializes in real time, using a method known as Bernoulli sampling (Sarndal et al., 1992).

In Bernoulli sampling, an inclusion probability, $p_h$, is assigned in advance to each stratum. At the midpoint of each sampling interval, a data logger reads the stage height and direction, retrieves the probability for the current stratum, and generates a uniform [0,1) random number, $c_i$. If $c_i < p_h$, the sampling unit is to be included in the sample, and the data logger activates a pumping sampler. With this method of sampling, the resulting stratum sample sizes are random variables.

In contrast to time strata, whose durations are selected in real time and are never combined, flow strata are defined before sampling begins and all partitions of the hydrograph (possibly including several runoff periods) in a given flow class contribute data to that stratum. Hence, the number of strata is the same regardless of the length of record (as long as the hydrograph 'visits' each flow class).

As with time-stratified sampling, estimates of totals and their variances can be obtained by applying the standard finite population formulae for stratified random sampling (Raj, 1968, p. 63). The load estimator is

$$\hat{Y} = \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi} \tag{1}$$

in which $L$ is the number of strata and, in stratum $h$, $N_h$ is the number of sampling units, $n_h$ is the realized sample size, and $y_{hi}$ is the $i$th sampled unit. The variance estimator is

$$\widehat{\mathrm{Var}}[\hat{Y}] = \sum_{h=1}^{L} N_h (N_h - n_h) \frac{s_h^2}{n_h} \tag{2}$$

where

$$s_h^2 = \frac{\sum_{i=1}^{n_h} \left( y_{hi} - \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi} \right)^2}{n_h - 1} \tag{3}$$

Because the stratum sample sizes are random, variances calculated in this way are conditional on the stratum sample sizes being what were actually obtained by chance.

In addition to estimating event loads, flow-stratified sampling can be used to compare mean loads within classes. Indeed, flow-stratified sampling grew out of attempts to develop 'flow-based standards' for water quality variates. Such standards associate flow classes with mean loads, which can be compared for different time periods and treatments to assess changes owing to management actions.

Using flow strata for comparisons, however, should be done with caution. With this method there is no untreated control area for reference. Unmeasured external changes could affect the sediment production of a watershed independently of the treatment (Stewart-Oaten et al., 1986). For example, a drought could cause reduction of vegetation in a watershed, temporarily changing the sediment-streamflow relationship. Without a reference watershed to detect such changes, measured differences could be mistakenly attributed to the treatment. The long times over which watershed studies are often conducted enhance this possibility.

### *2.5. Designing flow strata*

A flow-stratified sampling protocol for estimating event yields requires deciding on the number of strata, delineating stratum boundaries, setting the desired total sample size, and allocating the sampling effort among the strata. The protocol should depend on the purposes of sampling. For example, if stratum mean concentrations are of primary interest, as would be the case for standards comparisons, stratum boundaries may be chosen to meet legal or policy requirements. For estimating of total load, stratum boundaries that minimize the variance of the total load should be chosen. We offer guidelines when the objective is estimating the total load, but too much effort should not be spent on obtaining ‘optimal’ values. For our data, variance estimates were not very sensitive to modest differences in these sampling parameters. The greatest reduction in variance is realized with the decision to use stratified sampling rather than in its fine tuning. A general guiding principle when establishing strata is to partition the population in such a way that variation is minimized within and maximized among strata.

In most sampling problems, more strata mean lower variance and more expensive data collection. With the automatic procedures used for flow-stratified sampling, however, the cost does not rise appreciably with increasing numbers of strata if the total sample size remains constant. The most important factors limiting the number of strata are likely to be restrictions on the total number of sampled units that can be processed and the need to avoid sample sizes of zero in all strata. A balance is needed between having strata narrow and numerous enough to reduce the variance, while controlling the probability ($P_z$) of obtaining one or more stratum sample sizes of zero. This probability is given by

$$P_z = 1 - \prod_{h=1}^{L} [1 - (1 - p_h)^{N_h}] \tag{4}$$

The number of strata is best chosen using a priori knowledge of the behavior of the stream site being monitored, a determination of the sample size permitted by the budget, and an estimate of $P_z$. Wide strata can be split after data collection, but narrow strata cannot be combined after data collection unless the inclusion probabilities were the same in each. For streams in which most of the variation in sediment flux (discharge rate) occurs on the rising limb, that portion of the hydrograph should have the most strata.

A method for optimizing stratum boundaries was presented by Cochran (1963, pp. 128-133) and can be applied using available flow data and a sediment rating curve. The flow data are ordered and broken into a large number of classes (e.g. 500) with equal ranges in predicted sediment. The square roots of the class frequencies are then accumulated. The optimal stratum boundaries for any desired number of strata are found at equal intervals on the cumulative square root scale. The procedure can easily be applied separately to rising and falling flow data if so desired. In that case, two sediment rating curves should be used.

Allocation of samples to strata should depend on the expected sizes and variances of the strata and the costs of sampling in each stratum. Because the cost of selecting a unit is essentially the same in any stratum, the cost information need not be considered. Neyman allocation is widely used for standard stratified sampling when costs are equal in all strata (Cochran, 1963, p. 97). Here, the allocation cannot be set to exact optimality owing to the random sample sizes, but approximately optimal stratum probabilities can be assigned using the Neyman sample sizes. For stratum $h$ and a desired total sample size of $n$, the 'Neyman inclusion probability', $p_h^*$, is computed as

$$p_h^* = \frac{n_h^*}{N_h} = \frac{nS_h}{\sum_{h=1}^{L} N_h S_h} \tag{5}$$

in which

$$S_h = \left[ \frac{\sum_{i=1}^{N_h} \left( Y_{hi} - \frac{1}{N_h} \sum_{i=1}^{N_h} Y_{hi} \right)^2}{N_h - 1} \right]^{1/2} \tag{6}$$

where, in stratum $h$, $n_h^*$ is the Neyman sample size, $S_h$ is the standard deviation, and $Y_{hi}$ is the $i$th sampling unit. Allocations for the comparisons described in the next section were based on known population values from a particular period; in practice, previous samples are needed to estimate values of $N_h$ and $S_h$.

It is particularly important to obtain reliable values of $N_h$ from flow duration data, so that $P_z$ can be calculated. If $P_z$ is too large, then it will be necessary to increase the inclusion probabilities above the Neyman values for strata with small expected sample sizes ($E[n_h] = N_h P_h$). If good sediment data are not initially available to estimate $S_h$, the avoidance of zero stratum sample sizes might become the primary allocation consideration. In that case, rather than using the Neyman allocation, the probabilities could be chosen to equalize expected sample sizes in all strata. When estimates of $S_h$ later become available, the allocation could be switched to Neyman.

## 3. Methods

### 3.1. Data collection

Streamflow, suspended sediment, and turbidity data were collected from 1982 to 1986 at a station on the 10442 ha North Fork of the Mad River in northern

California. Discharge and turbidity were measured continuously during storm runoff along with frequent concentration specimens pumped from a boom-mounted depth proportional intake (Eads and Thomas, 1983). A logarithmic turbidity-concentration relationship was developed that allowed prediction of concentration from turbidity (the regression had an $R^2$ of 0.988 with 52 observations) with negligible error.

Population units were taken as sediment loads delivered in 10 min intervals. Each unit yield was calculated from midpoint streamflow and turbidity-predicted concentration, applied to the entire 10 min period. Sampling was compared from these finite populations of 10 min loads.

We isolated several fragments of the 5 year record ranging in length from 3 to 31 days and calculated their sediment yields. Some of the periods were single storms (defined by the method of Hewlett and Hibbert (1967)) and others contained several storms. These fragments were used separately and in combinations to form varying-length records to compare performance of the three competing sampling schemes.

Three single-peak storms with peak discharges of 147, 60, and 32 m3 s-l were selected from the 1982-1983 record and designated as Events A, B, and C respectively. Three isolated periods of the 1982-1983 data (including Events A, B, and C) were linked to form a 39 day record with ten peaks above 15 $m^3$ $s^{-1}$ and a maximum discharge of 147 $m^3$ $s^{-1}$. Finally, a 92 day record was formed by combining the 39 day period with five additional isolated periods from the years 1984-1986. This period included 22 peaks above 15 $m^3$ $s^{-1}$, and a maximum discharge of 147 $m^3$ $s^{-1}$.

### 3.2. Comparisons

The three sampling designs were compared by their coefficients of variation, which we redefined in terms of totals instead of means. Our formulation expresses the standard deviation of the estimated total (sediment yield) as a percentage of the known total, eliminating the effect of storm size on the comparisons.

The SALT auxiliary variable was derived from a moderately large 1983 storm (peak discharge was approximately 103 $m^3$ $s^{-1}$), which was not used in any of the comparisons. The sediment rating curve for this storm has a ‘hysteresis loop’ with two fairly straight and nearly parallel point patterns corresponding to the rising and falling limbs of the hydrograph. These patterns were fitted separately after removing the points connecting the two straight portions of the data (Fig. 3). This method, which gave good prediction for the fitted storm (except for flows late in the rising limb), was then used for SALT sampling in all storms compared in the analysis.

Time-stratified sampling requires ‘sampling schedules’ that specify stratum length for different stages and stage conditions (assuming a standard two or three measurements per stratum) (Thomas and Lewis, 1993). Optimal schedules are difficult to create, but selecting progressively shorter strata (i.e. higher sampling rates) for higher and rising stages and longer strata for lower and falling ones works well. (Again, the main benefit comes from stratifying, not from optimizing stratum parameters.)

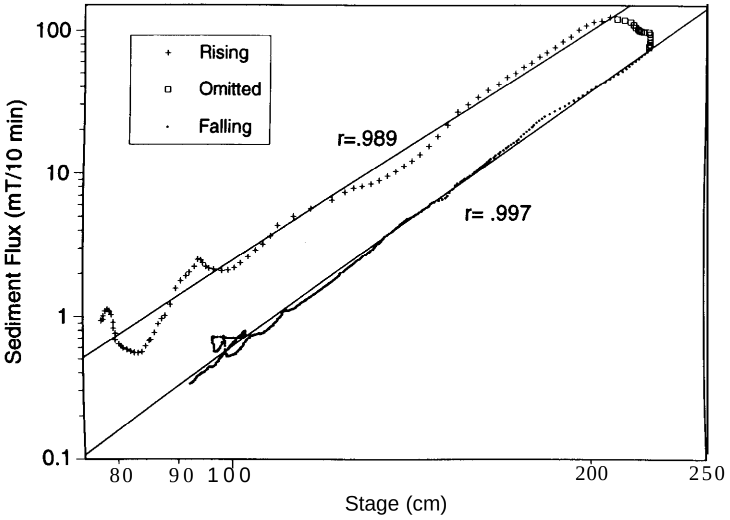

Fig. 3. Rising and falling stage regressions used for SALT auxiliary variable.

Varying the several stratum lengths while adhering to this general pattern yields different sample sizes (Table 1).

Time-stratified sampling requires long stratum lengths on low flows and during recessions to reduce sample size. If such a stratum extends into subsequent storm periods, its variance will be enormously increased, and the storm period may not be adequately sampled. Arbitrarily ending a stratum causes problems (such as what to do if no units have been sampled) that are hard to resolve. A technique that alleviates the problem is to combine the usual time strata with a flow stratum for the lowermost falling stage class. This stratum is ended and sampling shifted to a time-stratified

Table 1
Some time-stratified sampling schedules for real-time control of stratum lengths during storms

| Schedule no. | Stratum length (min) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Rising stage (cm) | | | | | Falling stage (cm) | | | |
| | $n_h$ | 0-122 | 122-152 | 152-183 | > 183 | >183 | 183-152 | 152-122 | 122-0 |
| 1 | 2 | 180 | 90 | 50 | 50 | 90 | 180 | 180 | 360 |
| 2 | 2 | 180 | 180 | 90 | 50 | 90 | 180 | 360 | 360 |
| 3 | 2 | 180 | 180 | 180 | 90 | 180 | 360 | 360 | 360 |
| 4 | 2 | 180 | 180 | 180 | 180 | 180 | 360 | 720 | 720 |
| 5 | 3 | 180 | 90 | 50 | 50 | 90 | 180 | 180 | 360 |
| 6 | 3 | 180 | 180 | 90 | 50 | 90 | 180 | 360 | 360 |
| 7 | 3 | 180 | 180 | 180 | 90 | 180 | 360 | 360 | 360 |
| 8 | 3 | 180 | 180 | 180 | 180 | 180 | 360 | 720 | 720 |

The stratum length is governed by stage height and direction. The number of sampling units, $n_h$, selected in each stratum is fixed at either two or three.

regime whenever the data logger senses a rise in stage. Subsequent data in that flow class continue to contribute to that stratum so it is unlikely to be empty for most records. We used this method to compute time-stratified estimates for the long-period comparisons, but not for the simple storms.

For the flow-stratified sampling design, variance reductions (computed from Eqs. (7) and (13) below) were modest using optimized boundaries or different numbers of strata, so we used four strata on each of the rising and falling limbs to match the time-stratified scheme with divisions at the same stages. The allocation among strata was optimized using the Neyman probabilities computed from a 31 day record in 1983 which contained five significant storms. None of these storms were used in the comparisons, because in an actual application, probabilities must be based on historical data.

Because complete populations were available it was possible to calculate the 'true' sampling variance for each plan without using sample estimates. For flow-stratified sampling, we used a modified variance formula which includes the additional variation due to random stratum sample sizes. It expresses the variance for the population of samples from which totals can be estimated, i.e. which do not include any zero stratum sample sizes. Denoting the vector of stratum sample sizes by $\boldsymbol{n} = (n_1, n_2, \ldots, n_L)$, the true variance of $\hat{Y}$ is

$$\begin{aligned}\mathrm{Var}[\hat{Y}] &= E_n[\mathrm{Var}[\hat{Y} \mid \boldsymbol{n}]] \\ &= E_{\boldsymbol{n}}\left[\sum_{h=1}^{L} N_h(N_h - n_h)\frac{S_h^2}{n_h}\right] \\ &= \sum_{h=1}^{L} N_h S_h^2\left(N_h E_{n_h}\left[\frac{1}{n_h}\right] - 1\right)\end{aligned} \tag{7}$$

The stratum sample sizes follow a truncated binomial distribution with density function

$$f(n) = \begin{cases}\binom{N}{n} p^n q^{N-n}/(1-a), & \text{for } n = 1, 2, \ldots, N; \\ 0, & \text{for } n = 0,\end{cases} \tag{8}$$

where the parameter $p$ satisfies $0 \leq p \leq 1$, the parameter N ranges over the positive integers, $q = 1-p$ and $a = q^N$. The subscripts $h$ have been dropped for simplicity. The moment generating function of $f$ is

$$m(t) = (pe^t + q)^N/(1 - a) - a/(1 - a) \tag{9}$$

from which it follows that

$$E[n] = m'(0) = Np/(1-a) \tag{10}$$

and

$$\mathrm{Var}[n] = m''(0) - [m'(0)]^2 = (1 - p)\left(\frac{Np}{1-a}\right) - a\left(\frac{Np}{1-a}\right)^2 \tag{11}$$

$E[1/n]$ can be approximated using a second-order Taylor series expansion of $g(n) = 1/n$ about the point $Np/(1 - a)$:

$$\frac{1}{n} = \frac{1-a}{Np} - \left(\frac{1-a}{Np}\right)^2 \left(n - \frac{Np}{1-a}\right) + \frac{2(1-a)^3}{(Np)^3} \left(n - \frac{Np}{1-a}\right)^2 \frac{1}{2!} + \ldots \quad (12)$$

$$\begin{aligned} E\left[\frac{1}{n}\right] &\doteq \frac{1-a}{Np} - \left(\frac{1-a}{Np}\right)^2 E[n - E[n]] + \frac{(1-a)^3}{(Np)^3} \mathrm{Var}[n] \\ &= \frac{(1-a)^2}{Np} \left(1 + \frac{1-p}{Np}\right) \end{aligned} \quad (13)$$

After reinserting the stratum subscripts $h$ on each variable in Eq. (13), this expression for $E[1/n_h]$ may be substituted into Eq. (7) for each stratum to obtain the approximate true variance of $\hat{Y}$ under Bernoulli flow-stratified sampling. For expected stratum sample sizes of at least 1.5, this variance is slightly larger than the usual estimate based on fixed stratum sample sizes.

## 4. Results

Comparisons among the three schemes for Events A, B, and C are shown as plots of coefficients of variation against sample size (Figs. 4, 5 and 6, respectively). SALT and flow-stratified sampling are shown as lines; because their allocations do not change with sample size, their variances are a smooth function of sample size. In contrast, the time-stratified sampling variance can only be computed for specific schedules, each of which has a unique allocation of the sampling effort. Hence, one point is shown for each time-stratified schedule.

Time-stratified sampling gave the lowest coefficient of variation for Events B and C across the range of sample sizes and was as low as or lower than the other schemes for

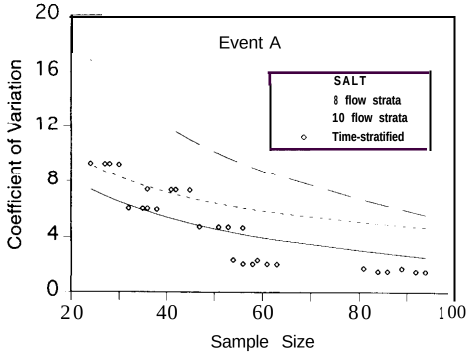

Fig. 4. Coefficients of variation over a range in sample size for SALT, time-stratified, and flow-stratified sampling methods applied to the simple hydrograph of Event A, which peaked at 147 $m^3 s^{-1}$. Flow-stratified sampling is shown with both eight and ten strata (the latter produced by splitting the upper flow classes).

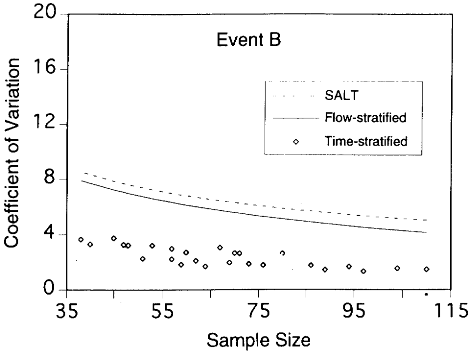

Fig. 5. Coefficients of variation over a range in sample size for SALT, time-stratified, and flow-stratified sampling (with eight flow strata) applied to the simple hydrograph of Event B, which peaked at 60 m$^3$ s$^{-1}$.

Event A. The basic time-stratified method was used (i.e. without a flow-stratum for the lowermost falling class) because, for single-peak events, there was no problem of subsequent storms beginning before a long falling stage stratum was completed. Flow-stratified and SALT sampling performed similarly, with two to three times the variation of time-stratified sampling for Events B and C. For Event A flow-stratified sampling with eight strata did not perform as well as SALT or time-stratified sampling. This poor performance is related to the larger size of Event A and is easily corrected as explained below.

The lower boundary for the highest rising and falling stage stratum was 1.8 m. Event A peaked at about 2.6 m, resulting in high variation in the two top strata (rising and falling) because of their wide range in stage. Because inclusion of each sampling unit is determined in real time as an independent Bernoulli trial, the strata can be 'split' at arbitrary stages after the data are collected as long as each resulting substratum has enough sampled units for estimation. Any flow stratum can be

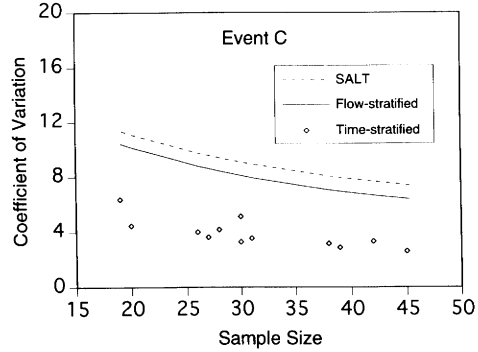

Fig. 6. Coefficients of variation over a range in sample size for SALT, time-stratified, and flow-stratified sampling (with eight flow strata) applied to the simple hydrograph of Event C, which peaked at 32 m$^3$ s$^{-1}$.

split, but, generally, the upper strata have the highest variance, so splitting them has the potential to provide the highest dividends.

Splitting strata is effective when it reduces the range (thus tending to reduce the variance) of sediment flux in the substrata. This is most likely for strata that are visited only once and in which sediment flux is highly correlated with stage. Fig. 7 shows sediment flux and stage in the 150–180 cm rising stratum for the 31 day period used to optimize flow-stratified sampling. This stratum was visited during two storms. Splitting the stratum at 165 cm does not reduce the range of sediment flux in the upper substratum. This is typical because relationships between flux and stage change substantially in different storms. In this case, the sum of the variances of the substrata totals is as large as the variance of the unsplit stratum total even though the range of the lower substratum was reduced. If we had been estimating individual storm loads the split would have greatly reduced the ranges of sediment flux for Storm 1 (in which sediment flux increases monotonically with stage), but not for Storm 2.

When the two upper strata were split for Storm A at 229 cm (making ten strata in all) the coefficient of variation was reduced by half or more (Fig. 4). Flow-stratified sampling with ten strata performed slightly better than time-stratified sampling for small sample sizes and not as well for larger samples. Events B and C did not occupy the highest strata so they were not split for these storms.

It is very difficult to predict individual storm flow durations with any confidence. Some strata may be visited only very briefly in a storm event, leading to significant probabilities of obtaining stratum sample sizes of zero. The probabilities ($P_z$) of one or more such occurrences in an event varied inversely with total sample size from 0.797 to 0.997 in Event A, from 0.006 to 0.362 in Event B, and from 0.003 to 0.136 in Event C. The higher probabilities for Event A result from the combination of low inclusion probabilities and shorter durations in the lowest flow classes.

The examples for storm events illustrate that the performance of flow-stratified sampling depends on sample size and the pattern of the particular storm for which load is being estimated. For smaller storms, flow-stratified sampling and SALT have

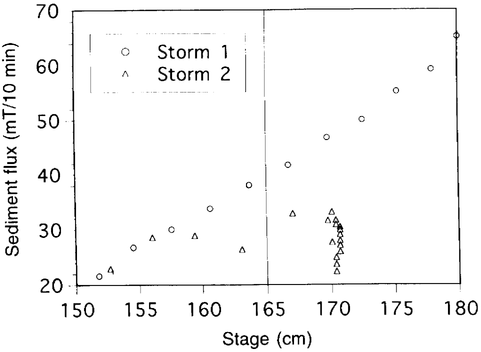

Fig. 7. Sediment flux vs. stage in a flow stratum containing data from two storm periods. Splitting the stratum at a stage of 165 cm lowers the variance of the total for Storm 1, but not for Storm 2 nor the hydrograph containing both storms.

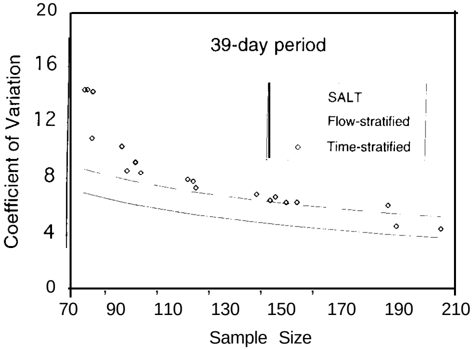

Fig. 8. Coefficients of variation over a range in sample size for SALT, time-stratified, and flow-stratified sampling (with ten flow strata) applied to the hydrograph of a 39 day composite period with ten peaks above 91 cm and maximum peak flow of 147 $m^3 s^{-1}$

comparable variance which is above that for time-stratified sampling. For large storms with hydrographs extending well into the upper strata (which are unbounded above), flow-stratified sampling (with stratum splitting) gives estimates about as good as time-stratified sampling, but in most cases only a partial storm can be estimated because there are no units sampled in the short-lived (lower) flow classes.

Applying the three schemes to the 39 day record produced a different pattern from that for the storm events (Fig. 8). Because of the large peaks in this period, the two uppermost flow strata were again split. The flow-stratified coefficients of variation are the lowest of any of the three methods. Time-stratified sampling and SALT performed comparably except that time-stratified sampling had relatively larger variance for smaller sample sizes.

The 92 day record further emphasized the better performance of flow-stratified sampling for longer complex hydrographs (Fig. 9). The time-stratified scheme performed poorly for smaller sample sizes and about as well as SALT for large samples. Splitting the two upper flow strata again greatly reduced the flow-stratified

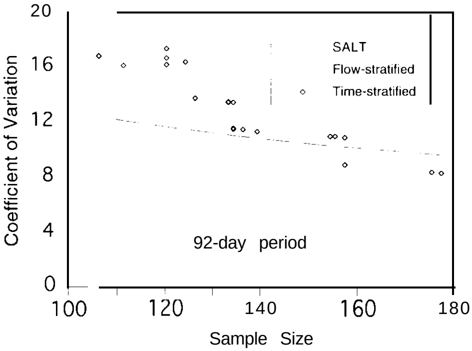

Fig. 9. Coefficients of variation over a range in sample size for SALT, time-stratified, and flow-stratified sampling (with ten flow strata) applied to the hydrograph of a 92 day composite period with 22 peaks above 91 cm and maximum peak flow of 147 $m^3 s^{-1}$

sampling variance, giving it the lowest coefficients of variation among these three schemes across the range of sample sizes.

## 5. Summary and discussion

Probability sampling of suspended sediment has important advantages even though it is more difficult to apply than deterministic methods. Validation is not needed at specific stations because the load and sampling variance estimates are known to be unbiased for any finite population. Variances under different sampling plans have different magnitudes, however, so they must be studied under different conditions to determine the best uses for each method. SALT, time-stratified, and flow-stratified sampling were compared by calculating true variances of load estimates from known sediment populations for simple and complex hydrographs.

Flow-stratified sampling is a form of stratified random sampling that is modified to use measurements of stage to govern the selection of population units in real time. Strata are formed for rising and falling stages, and probabilities reflecting expected stratum size and variation are assigned to each one. Concentration is determined from pumped samples for those population units for which a uniform [0,l) pseudo-random number (calculated in real time by a station data logger) is less than the stratum inclusion probability (Fig. 1). This method of sample selection is necessary because the population of flows is not known until the end of the period. However, this method of sample selection produces random stratum sample sizes, which raises the sampling variance slightly for estimates of total load.

The load for a monitored period may be estimated by the usual stratified random sampling formula for a population total. Because stratum sample sizes are random, the variance estimate is conditional on the actual sample size obtained.

In addition to estimating loads for specific events such as storms or seasons, flow-stratified sampling can be used for developing flow-based standards for assessing treatment effects. Stratum mean concentrations can be compared over selected periods to determine if they have changed. Flow strata should be used for standards only when a control watershed is monitored to insure that no external changes have occurred that affect sediment production.

Hydrographs for large storms may extend well into the upper strata, which are unbounded above. Under certain conditions, the high variance in these strata can be reduced by splitting them after the sample is collected-the only restriction being that each substratum have enough measurements to allow estimation. Variance reduction is most likely to be effective when a stratum was visited only once because, in this case, it is common for sediment flux to be highly correlated with stage.

The number of flow strata, their boundaries, and the inclusion probabilities must be assigned before sampling begins. These decisions depend on the purpose of the sampling program, the information accessible at the planning stages, and the resources available to operate the program. Sampling parameters can be altered as data are collected.

Time-stratified sampling tends to have lower variance than either flow-stratified or

SALT sampling for individual storm events (Figs. 4-6). For the large storm (Fig. 4) that extended well into the two upper strata, the flow-stratified sampling variance was substantially reduced by splitting the two upper strata into four substrata. However, the probabilities of obtaining zero stratum sample sizes are difficult to control and were unacceptably high in the largest storm event. Thomas and Lewis (1993) pointed out that SALT sample sizes are also difficult to control for storm events, resulting in very large samples in large storms, and very small samples with high variances in small storms. Time-stratified sample sizes are less variable and insure precise estimation of small as well as large events.

Flow-stratified sampling had lower variance than time-stratified or SALT sampling for longer hydrographs composed of multiple storms (Figs. 8 and 9). Time-stratified sampling gave the highest variance, except with large sample sizes. The poor performance of time-stratified sampling reflects the fact that a large number of time strata are needed to sample efficiently a lengthy hydrograph with multiple peaks; but this requires a large sample size, as variance calculation requires stratum sample sizes of at least two.

It is important to remember that the results of all the SALT comparisons in this paper assume a specific auxiliary variable based on stage. The auxiliary variable we used is based on much better data than would normally be available to develop a sediment rating curve, yet SALT was not the most efficient method in any of the comparisons. On the other hand, an auxiliary variable based on continuous measurements of turbidity might significantly improve SALT's performance in estimating suspended sediment loads. We could not use turbidity as an auxiliary variable in these comparisons because suspended sediment was computed directly from turbidity to construct our populations. These results can be summarized as follows.

(1) For estimating the total load of individual storms, time-stratified sampling gives the lowest variance.

(2) For estimating the total load of individual storms, both flow-stratified sampling and SALT have poor sample size control.

(3) For estimating the total load of long periods with many peaks, flow-stratified sampling gives the lowest variance.

(4) Flow-stratified sampling variance can often be reduced by splitting high-flow strata after data collection.

(5) Time-stratified sampling variance tends to increase more rapidly, relative to the other methods, as the sample size is reduced, especially in complex hydrographs.

(6) Although SALT was not the best method in any of these comparisons, its variance might be greatly reduced if a better auxiliary variable were available.

## Acknowledgments

We thank the Simpson Timber Company for its cooperation with the Forest Service, US Department of Agriculture, in collecting data at the test site.

## References

Bennett, J.B. and Sabol, G.V., 1973. Investigation of sediment transport curves constructed using periodic and aperiodic samples. In: Proceedings of the IAHR Symposium on River Mechanics, Vol. 2, 9-12 January 1973, Bangkok, Thailand, pp. 49-60.

Beschta, R.L., 1978. Long-term patterns of sediment production following road construction and logging in the Oregon coast range. Water Resour. Res., 14(6): 1011-1016.

Campbell, F.B. and Bauder, H.A., 1940. A rating-curve method for determining silt-discharge of streams. EOS (Trans. Am. Geophys. Union), 21: 603-607.

Cochran, W.G., 1963. Sampling Techniques, 2nd edn. Wiley, New York, 413 pp.

Cohn, T.A., Caulder, D.L., Gilroy, E.J., Zynjuk, L.D. and Summers, R.M., 1992. The validity of a simple statistical model for estimating fluvial constituent loads: an empirical study involving nutrient loads entering Chesapeake Bay. Water Resour. Res., 28(9): 2353-2363.

Colby, B.R., 1956. Relation of sediment discharge to streamflow. US Geol. Surv. Open File Rep.

Eads, R.E. and Thomas, R.B., 1983. Evaluation of a depth proportional intake device for automatic pumping samplers. Water Resour. Bull., 19(2): 289-292.

Ferguson, R.I., 1986. River loads underestimated by rating curves. Water Resour. Res., 22(l): 74–76.

Ferguson, R.I., 1987. Accuracy and precision of methods for estimating river loads. Earth Surf. Processes Landforms, 12: 95–104.

Gilroy, E.J., Hirsch, R.M. and Cohn, T.A., 1990. Mean square error of regression-based constituent transport estimates. Water Resour. Res., 26(9): 2069-2077.

Glysson, G.D., 1989. Sediment transport curves. US Geol. Surv. Open File Rep., 87-218, 47 pp.

Gurnell, A.M. and Fenn, C.R., 1984. Box-Jenkins transfer function models applied to suspended sediment concentration–discharge relationships in a proglacial steam. Arct. Alp. Res., 16(l): 93–106.

Guy, H.P. and Norman, V.W., 1970. Field methods for measurement of fluvial sediment. In: Applications of Hydraulics, Book 3, Chapter C2, Techniques of Water Resources Investigation of the US Geological Survey, US Government Printing Office, Washington, DC, 59 pp.

Hansen, M.H. and Hurwitz, W.N., 1943. On the theory of sampling from finite populations. Ann. Math. Stat., 14: 333-362.

Hewlett, J.D. and Hibbert, A.R., 1967. Factors affecting the response of small watersheds to precipitation in humid areas. In: W.E. Sopper and H.W. Lull (Editors), Forest Hydrology. Pergamon, New York, pp. 275-290.

Koch, R.W. and Smillie, G.M., 1986. Bias in hydrologic prediction using log-transformed models. Water Resour. Bull., 22(5): 717-723.

Lemke, K.A., 1991. Transfer function models of suspended sediment concentration. Water Resour. Res., 27(3): 293-305.

Miller, C.R., 1951. Analysis of flow-duration sediment-rating curve method of computing sediment yield. US Bureau of Reclamation Report, Denver, CO.

Norick, N.X., 1969. SALT (selection at list time) sampling: a new method and its application in forestry. M.A. Thesis, University of California, Berkeley, 50 pp.

Porterfield, G., 1972. Computation of fluvial-sediment discharge. In: Applications of Hydraulics, Book 3, Chapter C3, Techniques of Water Resources Investigation of the US Geological Survey, US Government Printing Office, Washington, DC, 66 pp.

Raj, D., 1968. Sampling Theory. McGraw-Hill, New York, 302 pp.

Rieger, W.A. and Olive, L.J., 1984. The behaviour of suspended sediment concentrations during storm events. In: R.J. Loughran (Compiler), Drainage Basin Erosion and Sedimentation. University of Newcastle, N.S.W.

Rieger, W.A. and Olive, L.J., 1988. Channel sediment loads: comparisons and estimation. In: R.F. Warner (Editor), Fluvial Geomorphology of Australia, Academic Press Australia, N.S.W.

Sarndal, C., Swennson, B. and Wretman, J., 1992. Model Assisted Survey Sampling. Springer, New York, 694 pp.

Stewart-Oaten, A., Murdoch, W.W. and Parker, K.R., 1986. Environmental impact assessment: ‘pseudo-replication’ in time? Ecology, 67(4): 929–940.

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

Thomas, R.B., 1988. Monitoring baseline suspended sediment in forested basins: the effects of sampling on suspended sediment rating curves. Hydrol. Sci. J., 33(5): 499-514.

Thomas, R.B., 1989. Piecewise SALT sampling for estimating suspended sediment yields. US For. Serv. Gen. Tech. Rep., PSW-114, 11 pp.

Thomas, R.B. and Lewis, J., 1993. A comparison of Selection at List Time and time-stratified sampling for estimating suspended sediment loads. Water Resour. Res., 29(4): 1247-1256.

US Interagency Committee on Water Resources, Subcommittee on Sedimentation, 1963. Determination of fluvial sediment discharge. Measurement and analysis of sediment loads in streams, Rep. 14. US Government Printing Office, Washington, DC, 151 pp.

Vanoni, V.A. (Editor), 1975. Sedimentation Engineering. American Society of Civil Engineers, New York, 745 pp.

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

Walling, D.E., 1977b. Limitations of the rating curve technique for estimating suspended sediment loads, with particular reference to British rivers. IAHS Publ., 122.

Walling, D.E. and Webb, B.W., 1988. The reliability of rating curve estimates of suspended sediment yield: some further comments. IAHS Publ., 174.