WATER RESOURCES RESEARCH, VOL. 32, NO. 7, PAGES 2299-2310; JULY 1996

Turbidity-controlled suspended sediment sampling for runoff-event load estimation

Jack Lewis
Pacific Southwest Research Station, U.S. Forest Service, U.S. Department of Agriculture, Arcata, California

Abstract. For estimating suspended sediment concentration (SSC) in rivers, turbidity is generally a much better predictor than water discharge. Although it is now possible to collect continuous turbidity data even at remote sites, sediment sampling and load estimation are still conventionally based on discharge. With frequent calibration the relation of turbidity to SSC could be used to estimate suspended loads more efficiently. In the proposed system a programmable data logger signals a pumping sampler to collect SSC specimens at specific turbidity thresholds. Sampling of dense field records of SSC and turbidity is simulated to investigate the feasibility and efficiency of turbidity-controlled sampling for estimating sediment loads during runoff events. Measurements of SSC and turbidity were collected at 10-min intervals from five storm events in a small mountainous watershed that exports predominantly fine sediment. In the simulations, samples containing a mean of 4 to 11 specimens, depending on storm magnitude, were selected from each storm's record, and event loads were estimated by predicting SSC from regressions on turbidity. Using simple linear regression, the five loads were estimated with root mean square errors between 1.9 and 7.7%, compared to errors of 8.8 to 23.2% for sediment rating curve estimates based on the same samples. An estimator for the variance of the load estimate is imprecise for small sample sizes and sensitive to violations in regression model assumptions. The sampling method has potential for estimating the load of any water quality constituent that has a better correlate, measurable in situ, than discharge.

Introduction

The transport rate of suspended sediment is of considerable interest when studying catchment hydrology and the impacts of land management. Because suspended sediment is a nonpoint pollutant whose concentration in natural streams varies rapidly and unpredictably, transport changes associated with human causes are usually difficult to demonstrate. The relation between instantaneous measurements of suspended sediment concentration (SSC) and water discharge is generally too variable to detect shifts in the relation over time. Perhaps the most effective technique for identifying changes is comparison of sediment loads from comparable periods between similar paired gaging stations before and after some treatment or disturbance is applied to one of the watersheds. Long-term seasonal or annual loads are less variable than individual runoff event loads, but it may require many years to amass enough values to detect statistically significant changes. The time required may even exceed the longevity of effects being studied. Discrete storm event loads are thus more useful, but accurate values are essential in order to avoid masking real effects with measurement error. Methods are needed to obtain accurate load estimates at reasonable cost, and the accuracy of the load estimates must be demonstrable.

Cohn [1995] summarizes recent advances in statistical methods for estimating sediment and nutrient transport. A number of these methods can provide measures of the uncertainty of the load estimate. The uncertainty may be gaged by the variance which arises from the random selection of sampling times or from a random element in a fitted model (such as a regression) for sediment concentration or flux. The random sampling methods can be implemented by means of a programmable data logger that senses stage, determines sampling times, and triggers an automatic pumping sampler. Most of these methods utilize time or discharge in some way as a covariate or auxiliary variable to improve precision.

For runoff event load estimation, time-stratified sampling [*Thomas and Lewis,* 1993] is the most suitable random sampling procedure [*Thomas and Lewis,* 1995]. This method varies the sampling frequency between time periods (strata) according to discharge. Time-stratified sampling has been shown to be more efficient for event load estimation in a 100-krn^2 rain-dominated mountainous basin than two discharge-driven random sampling methods: selection-at-list-time sampling [*Thomas,* 1985] and flow-stratified sampling [*Thomas and Lewis,* 1995]. In time-stratified sampling, SSC specimens should be frequent relative to the duration of sediment pulses. Sample sizes of 20 specimens for events lasting up to 3 days generally achieve standard errors less than 10% of the load. However, it can be costly to monitor extended high-flow periods that produce larger sample sizes.

Because high-frequency sampling for SSC is often impractical and expensive, easier-to-measure surrogate variables are often monitored with in situ sensing devices [*Gilvear and Petts,* 1985; *Hasholt,* 1992; *Jansson,* 1992; *Lawler et al.,* 1992]. The detailed records available from such studies can more than compensate for imperfect relations between the surrogate variables and SSC. Most devices measure the attenuation or scattering of an incident beam of radiation. Attenuance turbidime-

Paper number 96WR00991.

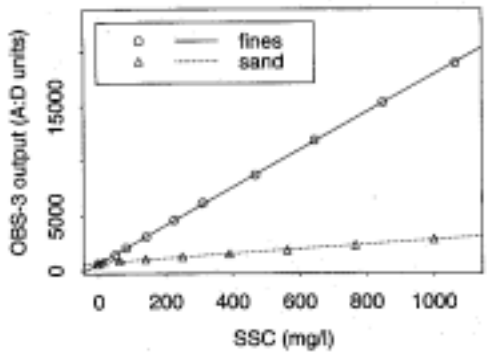

Figure 1. Infrared backscatter turbidity meter responds linearly to sediments of a given size distribution, but is much more sensitive to fines than to sand size material.

ters measure the loss in intensity of a narrow parallel beam or dual beams. Nephelometric turbidimeters measure light scattered at an angle (commonly 90° or 180°) to the beam and have been adopted by Standard Methods as the preferred method of turbidity measurement [*American Public Health Association* (APHA), 1985, p. 134]. Turbidimeter response to a given suspension is governed mainly by the light source, detector, and optical geometry. With sensors calibrated to give a linear response to standards, the response to varying SSC should be linear if the physical properties of the suspended particles are constant [*Gippel,* 1995].

We tested an OBS-3 (D and A Instrument Company) (trade names are used for information only and do not constitute an endorsement by the author or the U.S. Department of Agriculture) nephelometric turbidimeter that measures infrared backscatter at 180° with two particle size classes, 0-63 μm and 63-125 μm, sieved from field samples. Suspensions were spun on a magnetic stirring plate. Readings were taken once per second and averaged over 60 s. The response to fines was linear (r^2 = 0.9999) and to fine sands was nearly linear (r^2 = 0.995, with a quadratic fit r^2 = 0.9997). Sensitivity to the fines was much greater than to sands (Figure 1). Qualitatively similar results were reported by *Foster et al.* [1992] for five particle size bands of in-stream sediments with median diameters ranging from 4 to 63 gm. Explained variances exceeded 98.4% for all sizes and regression slopes varied by a factor of 25 between particle sizes. Therefore, in a natural stream it will be difficult to detect changes in total SSC unless they are associated with changes in the concentration of fines.

In addition to the sensor design and particle size distribution, turbidity is affected by particle shape, composition, and water color [*Gippel,* 1989]. For the same concentration and particle size, organic particles can give attenuance turbidity values two to three times higher than mineral particles [*Gippel,* 1995]. Color-producing dissolved organic substances increase attenuance turbidity but reduce nephelometric turbidity. Infrared turbidimeters are unaffected by water color, but are less sensitive than visible light turbidimeters to scattering from fines.

Gippel [1995] cites many studies documenting temporal variations in suspended solids that confound turbidity measurements in rivers. Variations in particle size can occur seasonally [*Bogen,* 1992] and during storm events [*Bogen,* 1992; *Peart and Walling,* 1992]. Although particle sizes can remain nearly constant [*Fleming and Poodle;* 1970], they frequently increase with concentration [*Frostick et al.,* 1983; *Long and Qian,* 1986; *Reid and Frostick,* 1987], and have also been observed to decrease with concentration [Colby *and Hembree,* 19551. Some riverine systems have an unstable relation between particle size distribution and discharge [*Walling and Moorehead,* 1987]. Particle mineralogy of transported sediments can change as a result of variable source areas [*Weaver,* 1967; *Richards,* 1984; *Johnson and Kelley,* 1984] and shifts between base flow and storm flow [*Wall and Wilding,* 1976]. The amount and types of organic material can also be expected to vary both seasonally and during storm runoff [*Ongley et al.,* 1982; *Walling and Kane,* 1982; *Hadley et al.,* 1985]. Water color variations owing to the presence of dissolved organics tend to be related to discharge, but they may be poorly related to suspended particles since their source areas often differ [*Gippel,* 1987].

Despite these complications, *Gippel* [1995] states that adequate relations between field turbidity and sediment concentration can be expected in most situations. Particle size variations are generally small or associated with variations in concentration. Turbidity data should be able to improve estimates based on infrequent measurements of concentration. However, in the presence of so many confounding factors, turbidity should not be used as a substitute for sediment concentration without careful study of the relation between turbidity and suspended load for any proposed monitoring sites. Without accompanying concentration data, there is no assurance in the quality of the estimate's.

Turbidity could be used more effectively than discharge as an auxiliary variable in selection-at-list-time sampling [*Thomas, 1985]* or stratified random sampling. The difficulty in selection at list time is that with a good auxiliary variable, sample sizes are approximately proportional to the sediment load, which can vary by orders of magnitude among estimation periods. Unpredictability of sample sizes is not as serious a problem in stratified sampling. Stratifying by turbidity should be at least as effective as flow stratification in estimating seasonal or annual loads. For estimating event loads, turbidity-stratified samples would likely be plagued with stratum sample sizes of 0 or 1, just as in flow stratification [*Thomas and Lewis,* 1995]. This paper tests a more direct application of turbidity for event load estimation.

By using weekly or biweekly calibration of turbidity with concentration data, *Krause and Ohm* [1984] were able to estimate loads in an estuary to within 12%. If the composition is stable enough, a few well-chosen SSC specimens may be all that are required to periodically recalibrate the relation. Using turbidity to activate a pumping sampler could automatically provide SSC specimens under certain conditions at specific turbidity levels to maintain a reliable relation. If hysteresis is present within a runoff event, as was observed, for example, by *Gilvear and Petts* [1985], a series of well-spaced specimens covering the range of the data could define the hysteretic loop and enable reliable load estimation. Further, the effects of instrument problems such as drift, temperature sensitivity, and minor optical fouling (e.g., by algal growth) are minimized by regular calibration.

The remainder of this paper will investigate, via simulation experiments, the feasibility and efficiency of turbidity-controlled sampling for suspended load estimation in a small Pacific coastal watershed. The best algorithms for sampling and runoff event load estimation are identified by application to very dense field records of concentration and turbidity.

Data Collection

Study Site

The Arfstein gaging station drains 384 ha of the Caspar Creek Experimental Watershed, a steep, rain-dominated forested catchment on the coast of northern California. The channel is 4 m wide, and the mean annual flood is about 2.3 $m^3 s^{-1}$. This gravel-bed stream typically transports about two thirds of its sediment load in suspension. The load is composed primarily of silts and clays from soils developed in sandstone and shale units of the Franciscan Assemblage [*Bailey et al.,* 1964]. Sand fractions in the suspended load generally increase with SSC and occasionally exceed 50% at high concentrations (Figure 2).

For both study phases described below, water specimens were pumped during storm events at 10-min intervals by ISCO model 3700 pumping samplers from an intake nozzle mounted on a boom designed to automatically position the intake at 60% of the flow depth [*Eads and Thomas,* 1983]. The SSCs from depth-integrated specimens agree well with those from simultaneous pumped specimens (Figure 3). Three pumping samplers, holding 24 bottles each, were filled in rotation. Bottles were filled to a volume of 250-450 mL. Initiation and cessation of sampling were controlled by an ONSET data logger programmed in BASIC.

Phase 1 (1991-1993)

This preliminary phase provided turbidity at 10-min intervals, with measurements of SSC on a substantial subset. Storm event sampling was initiated automatically by the data logger upon sensing a combination of a specified rainfall intensity (from a tipping-bucket rain gage at the site) and a minimum increase in water depth. Turbidity was measured in the laboratory with a Fisher model DRT 1000 turbidity meter. For each storm, specimens were divided into 10 turbidity classes whose boundaries were at equal intervals on a logarithmic scale. A random subset of specimens was selected from within each class for measurement of SSC. The number of specimens analyzed for SSC was 24 in storm events lasting up to 99 intervals (990 min), and one fourth the number of intervals in longer events. Additional specimens were analyzed on two occasions to obtain more continuous data. SSC was measured by standard gravimetric analysis using vacuum filtration through 1-μm filters. A total of 434 concentrations, representing five complete storm events and three partial events, were measured.

The relation between SSC and turbidity indeed varies by storm (Figure 4). In storms 91-2, 92-1, and 93-3 the relation is linear; in 93-1, 93-2, and 93-10 it is curvilinear. In 91-2, 92-1, and 93-1 hysteresis is evident (higher SSC earlier for given turbidity), but there is no hysteresis in 93-2 or 93-3. When plotted together (Figure 5a), the individual storms are segregated. To examine the effect of applying an external relation (based on storms other than that being estimated), sediment loads for each storm were calculated using both a customized fit (Figure 4) and an overall fit based on all the data except those from the storm being estimated. The overall fits were linear in the cube roots of SSC and turbidity. *Duan's* [1983] smearing estimator was used to correct for bias due to back transforming the SSC predictions. Although the overall fits were quite good, they were not always a good model for the excluded storm (Table 1). Differences between the estimates from the overall and customized fits varied from less than 1% to 250%, indicating that external relations are unreliable for predicting event loads. These results are optimistic, because in practice one rarely, if ever, has this quantity or range of data to determine historical relations, which then may be applied to more distant time periods. Of course, application of contemporary relations to historical periods is also subject to such errors.

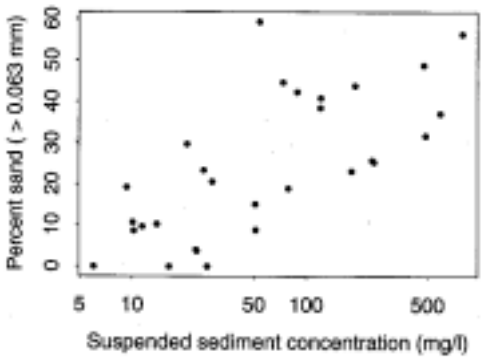

Figure 2. Relation of sand fraction to SSC for fixed-intake pumped SSC specimens collected at Arfstein station in water years 1986-1988.

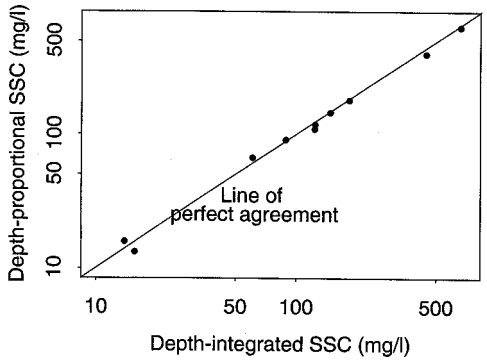

Figure 3. Relation of SSC between pumped specimens from depth-proportional intake boom and depth-integrated hand-drawn specimens from Arfstein station in water years 1991-1995.

Phase 2 (1994-1995)

Phase 2 provided both turbidity and SSC at 10-min intervals during the major portion of seven storm events, enabling calculation of "true" loads for simulation purposes. A secondary purpose of this phase was to test the feasibility of an in situ turbidity probe. Turbidity was recorded in real time using an Analite 156 (McVan Instruments, Pty., Limited) turbidimeter mounted in a protective housing on the boom near the pumping sampler intake. This nephelometric turbidimeter measures infrared 180° backscatter. Sampling was begun at the third interval above 20 formazine turbidity units (FTU) at the start of each storm; the time between pumped specimens was increased to 2 hours when turbidity later fell below 30 FTU for three intervals, and sampling was stopped at the third interval below 20 FTU. All pumped specimens were analyzed in the laboratory for SSC. A total of 1054 SSC/turbidity pairs in seven storms were measured. At the end of phase 2, one of the three pumping sampler intake hoses was found partially blocked with sediment. A plot of SSC against turbidity revealed a systematic reduction in SSC for the blocked intake; however the SSC was

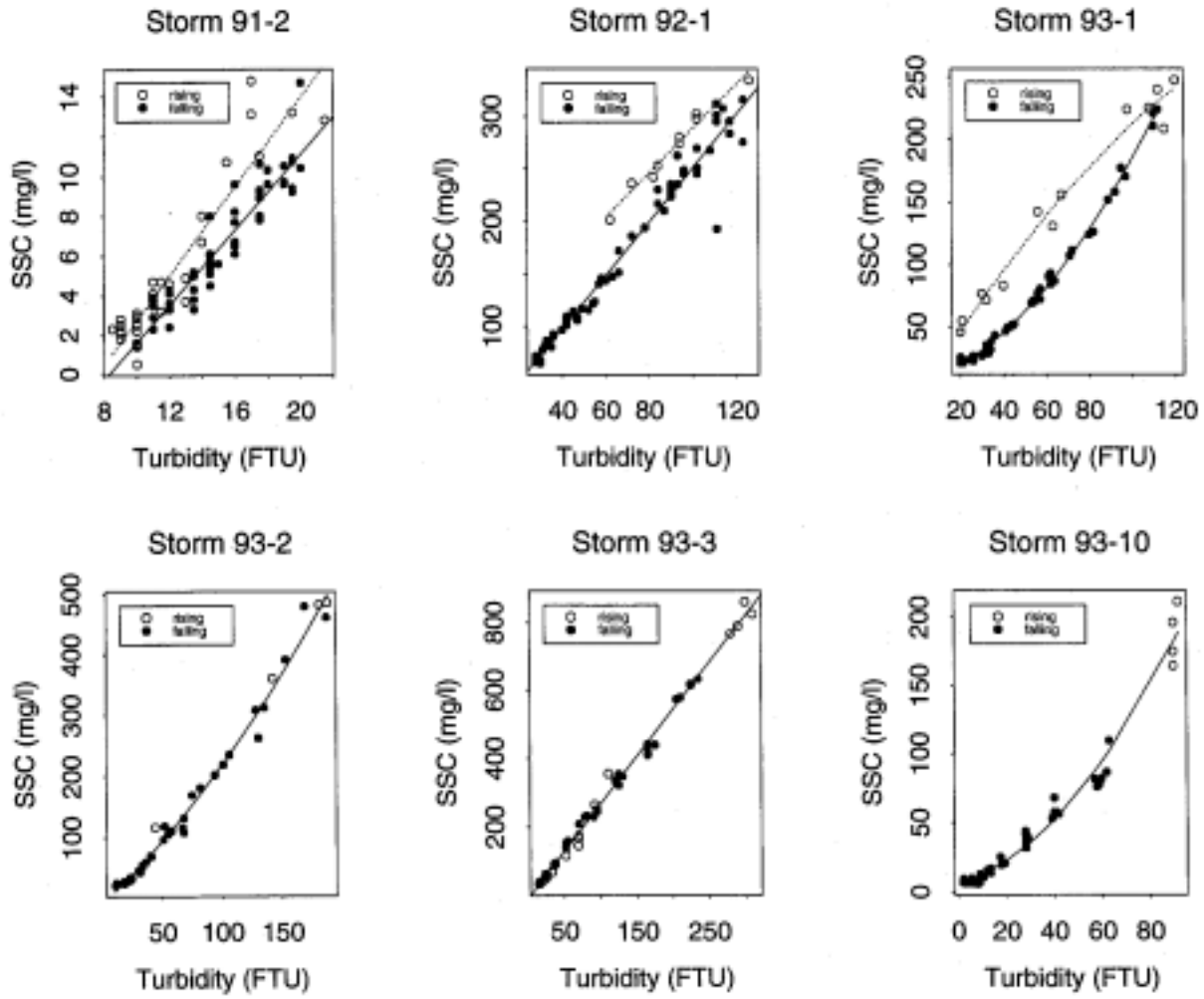

Figure 4. Relation of SSC to turbidity for phase 1 storms varies in form. Separate relations are sometimes apparent for rising and falling turbidities.

actually better correlated with turbidity than that from the other two samplers. Apparently, only some of the coarse sediments were being excluded. The SSCs from this sampler were adjusted using a linear transformation that brought the regression coefficients into agreement with that of the other samplers. .

As in phase 1, SSC is well correlated with turbidity, but individual storms are again segregated (Figure 5b). The five largest storms are shown in Figures 6 and 9. Storms 95-1a and 95-1b are two peaks from a single rainfall event, separated by 11 hours of missing data. Except for storms 95-2 and 95-6, the SSC turbidity plots appear to be linear. Hysteresis is present in 95-1a and, to a lesser degree, in 94-5 and 95-1b. The sediment loads for phase 2 are displayed in Table 2. The loads were computed by summing the products of discharge and SSC, which were assumed constant for each 10-min interval. An average of 75% of the sediment transport occurred during falling turbidities because of the lengthy recession periods. For comparison with phase 1, the sediment loads were also predicted using *Cohn et al.'s* [1989] minimum variance unbiased estimator from log linear models based on all the data except those from the storm being estimated. Prediction errors varied from -29% to 62%, once again confirming the need to calibrate the relation between turbidity and SSC for individual storms. It thus becomes useful to develop a protocol for collecting SSC specimens that will provide adequate calibrations for individual storms as efficiently as possible.

Simulations

Sampling Protocol

The calibration sample (set of SSC specimens) for a storm should cover the range in turbidity and should include any major swings in turbidity because they might reflect calibration shifts. One method of collecting such a sample is to establish turbidity thresholds for sampling. A programmable data logger, sensing that a threshold has been reached, sends a signal to an automatic pumping sampler to fill a bottle. Because more sediment is discharged while turbidity is in a recession mode, more thresholds are needed for falling turbidities than rising ones. Thresholds also need to be scaled to limit sample sizes while sampling arbitrary storms whose loads can span several orders of magnitude. A uniform sampling protocol that adequately defines loads for small storms could tremendously oversample large storms. One strategy is to scale the thresholds so that their density decreases with increasing turbidities. This can be accomplished by establishing thresholds that are uniformly spaced after transformation by logarithms or by a power

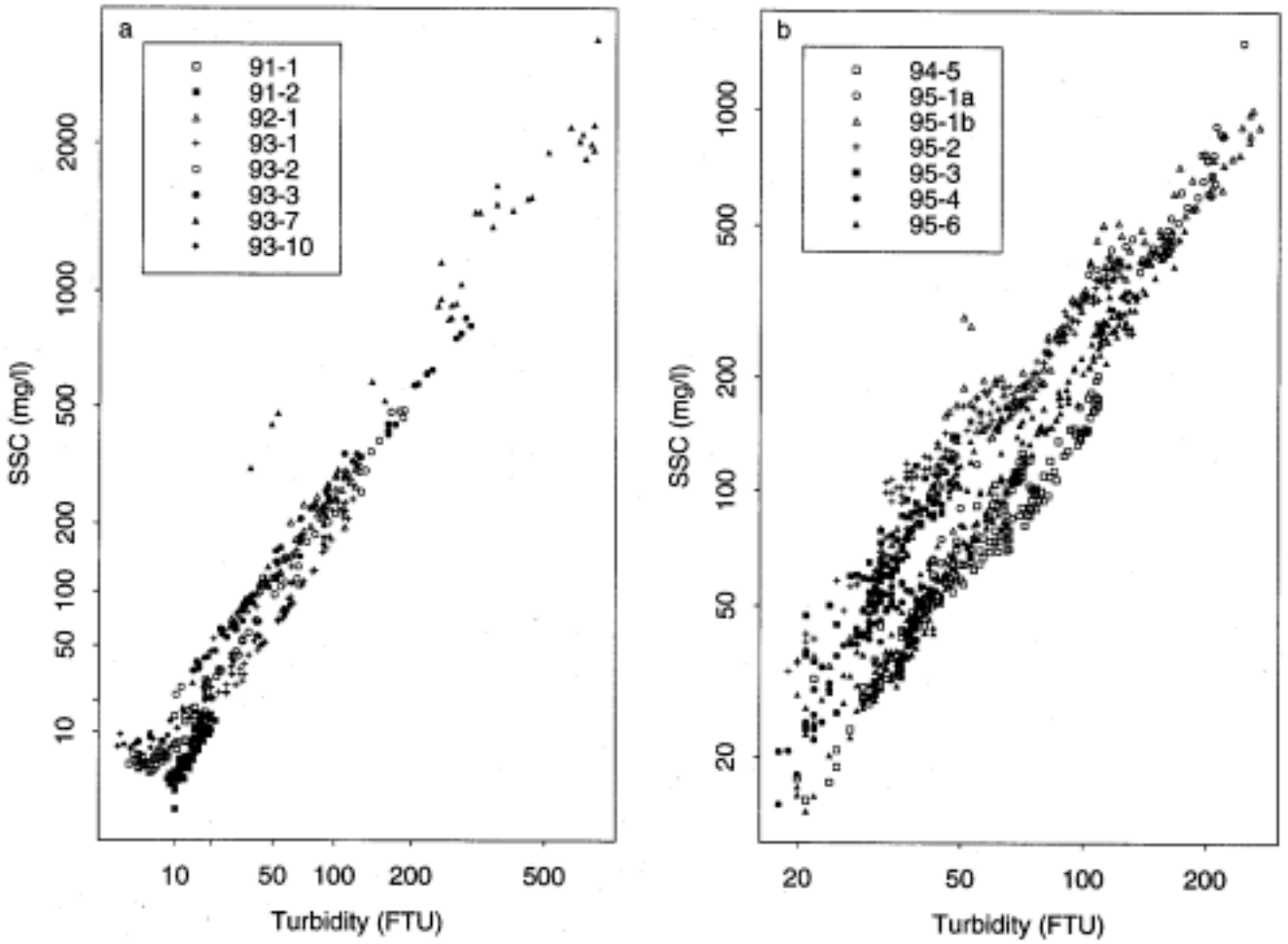

Figure 5. Scatterplots of SSC versus turbidity were segregated according to storm in both (a) phase 1 and (b) phase 2. To linearize the plots, axis scaling is cube root in Figure 5a and logarithmic in Figure 5b.

between 0 and 1. To avoid sampling of ephemeral turbidity spikes that may be caused by passing debris, we require a threshold to be met for two intervals before signalling the sampler.

A rule is also needed to reliably detect changes between rising and falling turbidity conditions so that the correct set of thresholds may be invoked. To detect a reversal, we require the turbidity to drop a specified amount below the preceding peak, or rise a specified amount above the preceding trough. For Caspar Creek we have specified 10% of the prior peak or 20% of the prior trough, but at least 5 FTU in all cases. Two different values are needed because turbidity spikes are mostly positive. As an additional precaution against false reversals, we require the turbidity to drop for at least two intervals before declaring a shift to recession and to rise for at least two intervals before declaring a new rise. At the time a reversal is detected a specimen is collected if a threshold has been crossed since the preceding peak or trough, unless another reversal occurred within the last hour. This set of rules was found to provide reasonable assurance of collecting a pumped specimen as soon as possible after a reversal while avoiding extraneous specimens in the presence of a fluctuating turbidigraph. For other streams it would be prudent to experiment, as we did, with a continuous turbidity record before deciding on a particular protocol.

Table 1. Sediment Loads for Phase 1

Storm	Estimated Load, kg: Individual	Estimated Load, kg: Overall	Difference, % of Individual
91-1	299	412	37.9
91-2	446	1564	250.9
93-1	6882	9934	44.3
93-2	19034	21224	11.5
93-3	90122	90574	0.5
93-10	10978	13936	26.9

Loads are based on individual storm fits and overall fits of $SSC^{1/3}$ to $turbidity^{1/3}$ for the combined data except that storm being estimated. Partial storms, 92-1 and 93-7, are not included.

Simulation Procedure

Storm events from phase 2 were sampled repeatedly with varying thresholds and fitting procedures to evaluate turbidity-controlled sampling and load estimation procedures. Five of the seven phase 2 storm events (Figure 6) reached turbidities of 100 FTU or more and were most suitable for sampling simulations. Events 95-3 and 95-4 were small storms reaching maximum turbidities less than 50 FTU. They produced inadequate sample sizes under any of the thresholds considered. Also, a large proportion (20% and 35% respectively) of the sediment from these two events was delivered during reduced

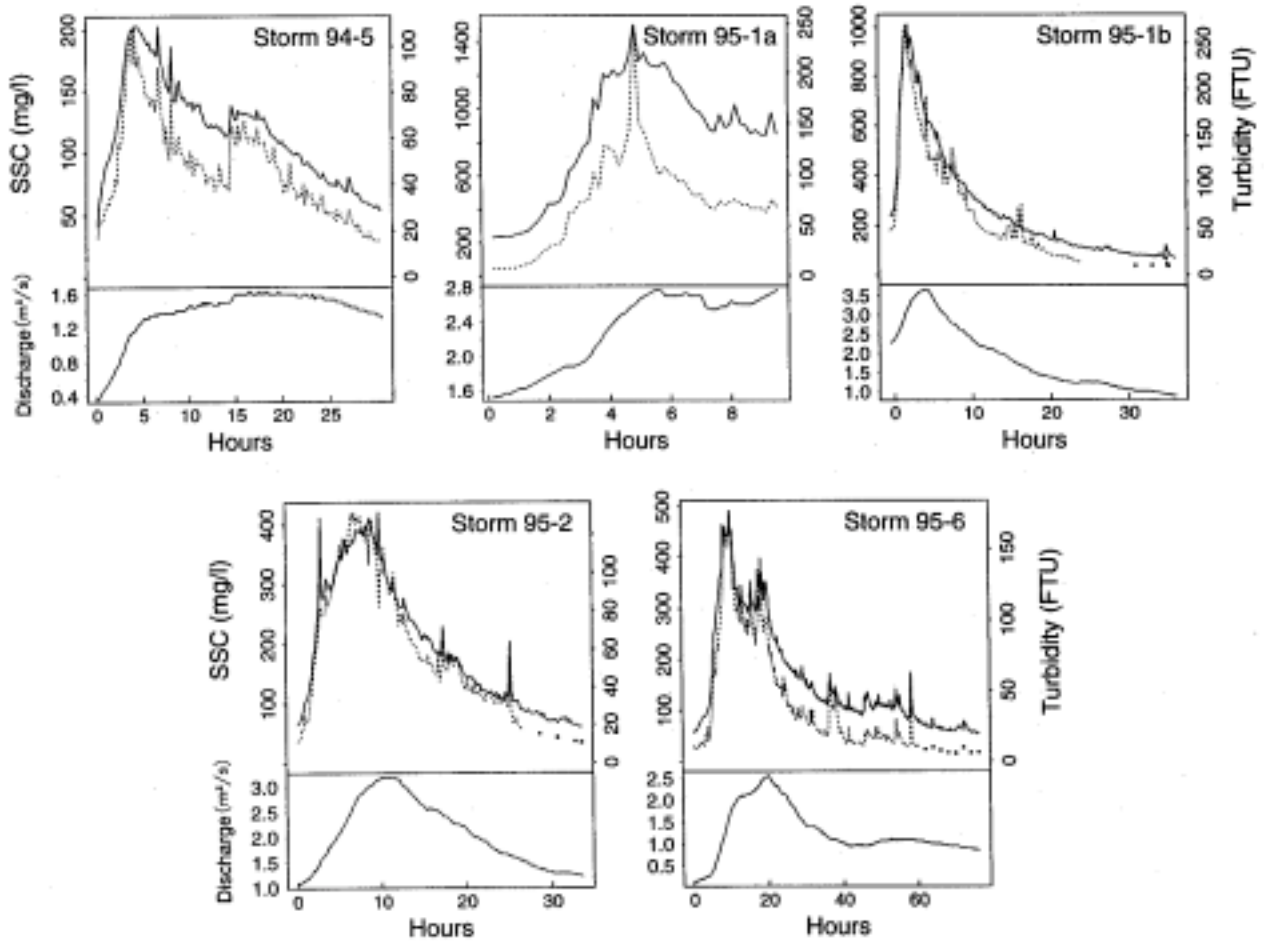

Figure 6. Phase *2* storms. Solid lines represent turbidity and discharge. Dotted line represents SSC.

sampling mode (when specimens were collected once every 2 hours) at receding turbidities under 30 FTU. Because they reflect a lower sampling intensity, data collected during reduced sampling mode in phase 2 storms were not included in the sampled populations. In the five storms used in the simulations, this decision resulted in the omission of an average of 3.5% and no more than 9% of the total sediment load in any storm.

Sampling variation was obtained by shifting the transformed threshold scales. A simulation produced 15 samples (sets of SSC specimens) per storm, one for each pair of scales (a rising and a falling set of thresholds). For example, starting with a set of FTU thresholds (21, 78, 171, and 300) spaced uniformly under a square root transformation, 15 rising scales of the form $\{j^2 , (j + d)^2 , \{j + 2d)^2 , (j + 3d)^2\}$ were obtained with $d = (300^{0.5} - 21^{0.5})/3$ and $j = i^{0.5}$, by letting i assume all integer values from 21 to 35. Falling scales were constructed analogously starting with eight values from 31 to 300 and $d = (300^{0.5} - 31^{0.5})/7$, letting i vary from 31 to 45. In practice, the scales would need to be extended above 300 to handle larger events. Square root, cube root, and logarithmic scales were considered. The mean sample sizes (number of SSC specimens per sample) for each scale type and storm are listed in Table 3.

Sediment loads were estimated by fitting SSC to turbidity for each simulated sample, using the fit to estimate SSC for all intervals, then summing the products of discharge and SSC for

Table 2. "True" Sediment Loads for Phase 2 Storms and Regression Estimates

Storm	True Load, kg	Estimate, kg	Error, % of True
94-5	12892	20832	61.6
95-1a	38088	35258	-7.4
95-1b	68543	59106	-13.8
95-2	47460	34431	-27.4
95-3	344	2447	-29.0
95-4	2202	1963	-10.9
95-6	46978	57888	23.2

"True" sediment loads are the sum of 10-min loads. Regression estimates are based on overall fits of log (SSC) to log (turbidity) for the combined data except that storm being estimated.

Table 3. Mean Sample Sizes From Application of 15 Pairs of Threshold Scales to Phase 2 Storms

	Storm				
Scale Type	94-5	95-la	95-1b	95-2	95-6
Square root	5.5	4.1	8.6	7.3	10.7
Cube root	5.9	4.1	8.5	7.5	11.4
Natural log	6.5	3.8	8.0	8.2	12.9

A pair of scales consisted of four rising turbidity thresholds and eight falling thresholds.

Table 4. Simulation Summaries

Simulation	Threshold Scale	Dependent Variable	Independent Variable	Order	Fits	Estimator	Storm, % RMSE 94-5	95-1a	95-l	95-2	95-6	Mean Rank
1	sqrt	c	t	1	1	ls	2.4	7.7	4.1	5.9	1.9	8.4
2	sqrt	c	t	2	1	ls	2.2	22.6	3.7	3.7	2.5	7.2
3	sqrt	c	t	1	2	ls	3.5	2.8	4.6	5.2	3.3	10.4
4	sqrt	c	t	2	2	ls	3.9	2.7	4.3	7.6	4.2	14.2
5	cbrt	c	t	1	1	ls	2.4	6.4	3.6	6.5	3.2	9.6
6	cbrt	c	t	2	1	ls	2.4	18.7	3.3	4.8	2.9	7.2
7	cbrt	c	t	1	2	ls	3.9	3.1	4.1	5.8	3.8	11.8
8	cbrt	c	t	2	2	ls	4.8	3.5	4.1	7.5	4.3	15.4
9	log	c	t	1	1	ls	3.7	9.3	3.4	6.9	3.0	12.4
10	log	c	t	2	1	ls	5.0	14.3	2.8	5.9	2.3	11.8
11	log	c	t	1	2	ls	4.1	8.7	3.4	6.8	3.3	13.2
12	log	c	t	2	2	ls	6.0	11.0	2.7	7.1	3.5	15.2
13	sqrt	sqrt(c)	sqrt(t)	1	1	sm	2.4	8.9	4.0	5.8	2.7	8.4
14	sqrt	sqrt(c)	sqrt(t)	1	2	sm	3.4	4.3	4.5	5.1	4.1	11.2
15	cbrt	cbrt(c)	cbrt(t)	1	1	sm	2.4	7.6	3.5	6.2	3.1	9.2
16	cbrt	cbrt(c)	cbrt(t)	1	2	sm	3.8	3.7	4.0	5.4	4.3	11.6
17	log	log(c)	log(t)	1	1	MVUE	3.1	9.1	4.6	6.3	3.0	13.8
18	log	log(c)	log(t)	1	2	MVUE	3.5	8.4	4.6	5.9	4.2	14.8
19	sqrt	log(c)	log(t)	1	1	MVUE	2.2	9.0	4.6	5.0	3.2	10.4
20	sqrt	log(c)	log(t)	1	2	MVUE	3.2	4.7	5.0	4.5	4.4	12.6
21	cbrt	log(c)	log(t)	1	1	MVUE	2.4	7.6	4.4	5.7	3.3	11.0
22	cbrt	log(c)	log(t)	1	2	MVUE	3.7	3.8	4.6	5.0	4.7	13.2
23	sqrt	log(c)	log(q)	1	1	MVUE	18.3	23.2	8.8	11.2	11.2	...
24	sqrt	log(c)	log(q)	1	2	MVUE	12.9	23.9	5.7	33.4	9.9	...

Each simulation generated 15 samples, one for each set of four rising turbidity thresholds and eight falling thresholds. sqrt, square root; cbrt, cube root; log, natural log; *c*, suspended sediment concentration; *t*, turbidity; *q*, discharge; order 1, linear; order 2, quadratic; ls, least squares estimate; sm smearing estimate; MVUE, minimum variance unbiased estimate. When number of fits is 2, separate curves were fit to rising and falling turbidities. In simulation 24, fits were applied to rising and falling discharges instead of turbidity when predictor was a function of *q*.

the storm. In half of the simulations a single curve was fitted to each sample (Table 4). In the other half, separate linear fits were automatically generated for rising and falling turbidities, unless either sample size was less than 2. Quadratic fits were generated for six of the simulations, and in 10 of the simulations, lines were fitted after transforming both variables, usually with transformations corresponding to the threshold scaling. Automated fitting procedures were adapted and each fitting method was examined for accuracy and consistency across storms.

For log linear fits, bias correction was carried out using *Cohn et al.'s* [1989] minimum variance unbiased estimator (MVUE). In most circumstances, mean square error for MVUE is very similar to that of *Duan's* [1983] smearing estimator [*Gilroy et al.*, 1990], but it is theoretically preferred because it is unbiased. MVUE is applicable only to log linear fits; therefore the smearing estimator was used for fits transformed by square roots and cube roots.

Results

The sampling/estimation methods were ranked according to the root mean square error (RMSE) computed from the 15 load estimates as a percentage of the known load for each storm. The mean rank for each method is listed in Table 4, alongside the percent RMSE for each storm. There was not a great deal of sensitivity to either the fitting method or the threshold scaling. Variable transformations tended to increase RMSE relative to the equivalent untransformed fits and, in general, errors increased in a progression from square root to cube root to logarithmic threshold scaling. The latter result might have been expected because samples collected on a square root scale are those most heavily weighted toward higher turbidities. The following paragraph refers to simulations 1 to 4, which did not incorporate variable transformations and which utilized square root threshold scaling. Similar patterns hold for the other scales.

A single linear equation fitted to the untransformed variables (simulation 1) resulted in RMSE varying from 1.9 to 7.7% for the five storms, with mean sample sizes between 4 and 11. Single quadratic fits (simulation 2) had lower mean ranks than linear fits, but sometimes failed dramatically (e.g., storm 95-1a), giving U-shaped curves and large extrapolation errors. When quadratic regression was successful, it was only slightly more accurate than linear regression except in storm 95-2 (Figure 7a), for which the RMSE was reduced from 5.9 to 3.7%. Separate linear fits for rising and falling turbidities (simulation 3) improved estimation for storm 95-1a, reducing the RMSE from 7.7 to 2.8%. If the rising and falling relations are similar, the larger combined sample usually covers a wider range of turbidity than the separate samples and generally results in a better fit than either individual fit. Because of their tendency to extrapolation errors, complex fitting procedures should be used with caution on small samples. In most cases a single linear fit performed nearly as well or better than other methods and may be preferred because of its consistency and parsimony.

The threshold scale-shifting algorithm did not give rise to estimates with independent errors because, frequently, samples generated by slightly offset scales shared some observations. To get more variation in the samples and to determine the effect of sample size, additional simulations were performed, among which the number of thresholds was varied from 2 to 8 on the rising limb and from 6 to 12 on the falling limb, that is, 4 more than on the rising limb. Regressions were

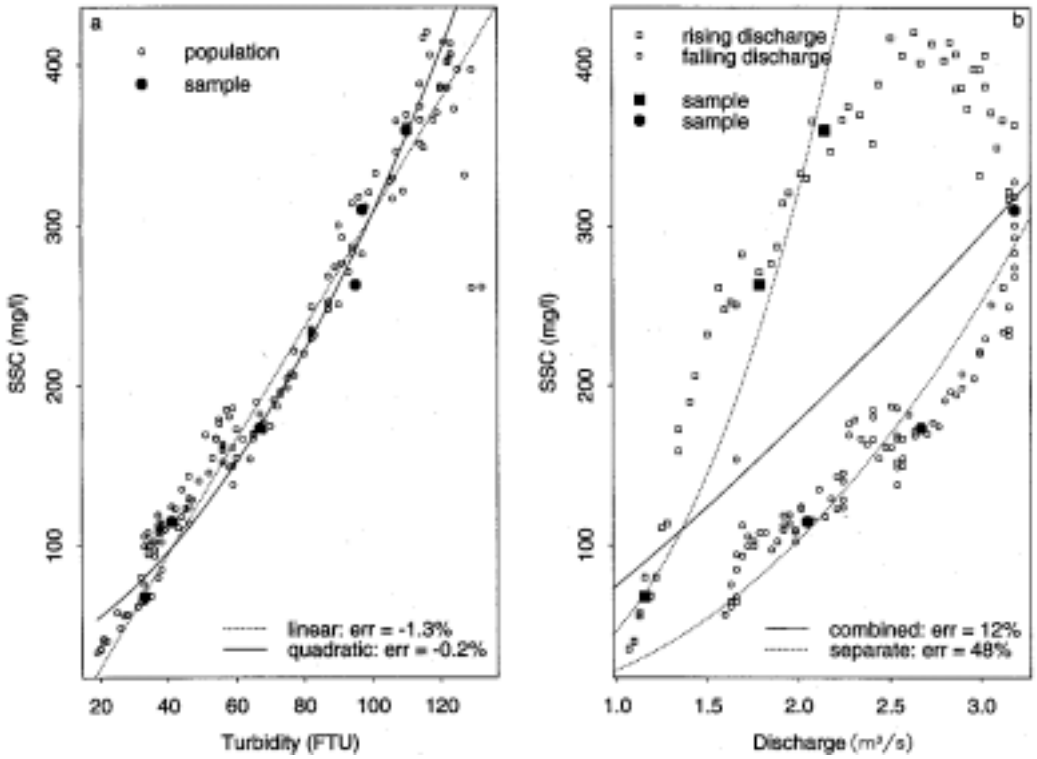

Figure 7. Example of simulation results from application of square root threshold scale to storm 95-2. (a) Linear and quadratic fits from simulations 1 and 2. (b) Log linear sediment rating curves applied first to all six points (simulation 23), then separately to three rising and three falling points (simulation 24). All four load estimates are based upon the same sample of six, but the errors (differences between estimated and true load as a percentage of the latter) based on the relations of SSC to turbidity are much smaller.

fit to untransformed sample data. Single linear fits were applied to three storms, single quadratic fits were applied to storm 95-2, and separate linear fits were applied to the rising and falling portions of storm 95-1a. The RMSE never exceeded 8% of the load for mean sample sizes of 3 or more and is seen to fluctuate or decline with sample size (Figure 8). For sample sizes of at least 5, RMSE is generally no more than 5% of the load.

To compare sediment rating curve estimation with turbidity-calibrated sampling, sediment rating curves (log linear fits of

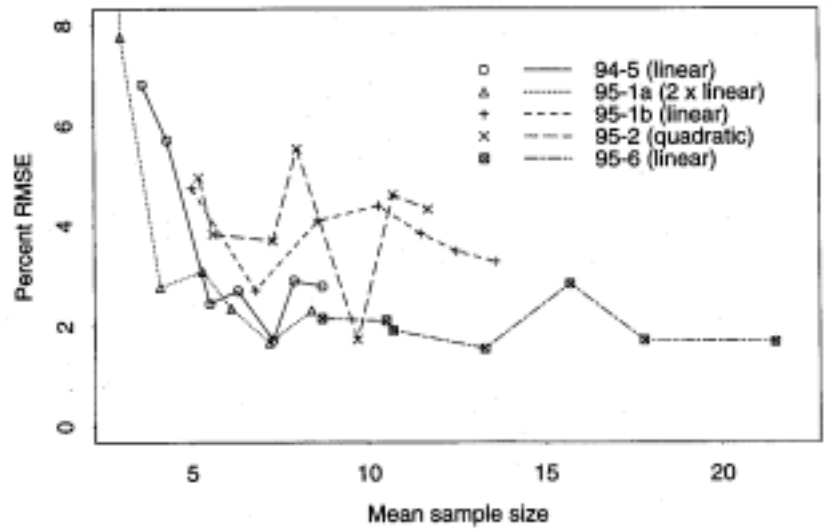

Figure 8. The number of simulated thresholds was varied on a square root scale and RMSE as a percent of the true load is displayed in relation to sample size. All load estimates are based on fits of untransformed SSC to turbidity. Linear fits were applied to all storms except 95-2. Fits were applied separately to data preceding and following the turbidity peak in storm 95-1a.

SSC to discharge) were fit to each of the samples generated by simulation 1 of Table 4. Loads were estimated with the MVUE estimator for single curve fits as well as for separate curves fit to the rising and falling portions. Rating curve estimates had RMSE 1.9 to 7.5 times larger than turbidity curve estimates when a single curve was fit, and 1.3 to 8.6 times larger with separate curves (Table 4, lines. 1 versus 23 and 3 versus 24). Simple log linear models are clearly inappropriate in the presence of such severe hysteresis (Figure 7b). Occasionally the automatically generated curves are nonsensical (e.g., decreasing), but, in practice, curves that appear reasonable might unfortunately be believed because of the absence of turbidity or SSC knowledge beyond the sample.

Variance Analysis

The variance of the load estimate can be estimated without bias if the regression model assumptions are satisfied. Suppose, for each interval *i*, the concentration (c_i) is modelled as a linear function of turbidity (t_i):

$$c_i = a + bt_i + \varepsilon_i \quad (1)$$

and ε_i are independent random errors, normally distributed with mean 0 and variance σ^2. Define the estimated load for interval *i* as

$$\hat{l}_i = kq_i\hat{c}_i \quad (2)$$

where q_i is the discharge at time *i*, c_i is the predicted concentration from the least squares estimates of *a* and *b* in the linear regression model, and *k* is a units conversion factor. The covariance of the estimated loads for two intervals *i* and *j* is given by

$$V_{ij} = \sigma^2 \mathbf{z}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{z}_j \quad (3)$$

where $\mathbf{X}$ is a matrix whose first column is all 1's and whose second column is the set of sample turbidities; $\mathbf{z}_i$ is the product $kq_i\mathbf{x}_i$, in which $\mathbf{x}_i = (1, t_i)$ is a predictor vector for an interval whose concentration is to be estimated. The load estimator is

$$\hat{L} = \sum \hat{l}_i \quad (4)$$

and the estimated variance of the total estimated load is the sum of the entries in the matrix $\mathbf{V} = (V_{ij})$:

$$\hat{V}_L = \sum\sum V_{ij} \quad (5)$$

In practice, σ^2 must be estimated as the mean square error from the regression. If the errors ε_i in the regression model are not independent and identically distributed, then this estimate is biased [*Seber*, 1977, p. 145]; consequently, the estimated variance of the load will be biased. In the methodology being explored in this paper, small samples of concentrations, too small to assess the error distributions in practice, are considered. It is reasonable to expect that the errors might not be independent because concentration specimens are obtained during the course of a highly correlated time series. In addition, one might expect errors to increase with increasing concentration.

To investigate the errors in variance estimation, linear models were fitted to simulated log linear data. The initial simulation was based on a sample of four turbidities (87, 77, 53, and 31) generated by iteration 1 of simulation 1, Table 4, for storm 94-5. The sampled turbidities were held constant while random concentrations were repeatedly generated from a log linear regression model fit to all the data from that storm. Note that the log linear curve is still quite linear within the range of the data (Figure 9), so it would not have been obvious in practice whether a linear or log linear model was more appropriate; nor would it have been apparent that the concentrations were log normally distributed for each turbidity. For each set of concentrations a linear model was fit to the data, the load was estimated from (1), (2), and (4), and the variance of the load was estimated from (3) and (5). *Cohn et al.'s* [1989] MVUE estimator for log linear models and its variance was also computed. Five thousand sets of concentrations were generated. With this sample size the simulated variance of the MVUE came within 0.4% of its true variance. The mean of the load estimates (L) agreed with the modelled mean load to within 0.10%. The percent RMSE for the load estimates was 6.5%. The mean of the variance estimates (V_L) exceeded the simulated variance of the load estimates by just 1.7%. Although the bias in variance estimation was small, the percent RMSE was 120% for the variance and 51% for the standard error, indicating that for this small a sample size the variance estimates are very poor. Note that in this case the apparent bias (3.0%) of the variance estimator for MVUE is larger than for the linear model. The MVUE variance estimator is slightly biased because, as in (3), σ^2 is unknown and must be estimated from the sample. The percent RMSE for the MVUE variance estimator (110%) and for the standard error (50%) were similar to that of the linear model. This simulation suggested that there is no practical disadvantage to using the simpler linear estimator even in the presence of lognormal errors, when the mean model is nearly linear.

The variance simulation was then repeated for each of the phase 2 storms, using the turbidity samples generated by iteration 1 of simulation 1, Table 4, for each storm. The sample sizes for the five storms are 4, 5, 9, 7, and 11. Table 5 summarizes these results, comparing the load and variance estimators for the untransformed model with that of a log linear model estimated using the MVUE. The maximum apparent bias in the load estimator was 4.01% and the maximum percent RMSE was 8.28%. The apparent bias in variance estimation from the untransformed model contrasted with the results for storm 94-5, varying up to 121%, with RMSE up to 244% for the variance and 72% for the standard error. Bias in both the load and variance estimators was greatest for those storms (95-la and 95-6) in which the postulated log linear model exhibited the most curvature (Figure 9). The MVUE load estimator, while unbiased, had RMSE values very similar to the untransformed models. The variance estimator for MVUE exhibited relatively small bias, up to 5.5%, with RMSE up to 98% for the variance and 43% for the standard error.

These simulations suggest that variance estimation can be improved by applying the correct model, but with the small sample sizes (4 to 11) being considered here, the uncertainty in variance estimation can still be quite large. The load estimates are very good in either case, which is comforting because it is difficult in practice to identify the correct model from a small sample.

Conclusions

Technology developed in the last decade provides opportunities to greatly improve suspended sediment transport estimation in streams and rivers. It is possible to collect essentially

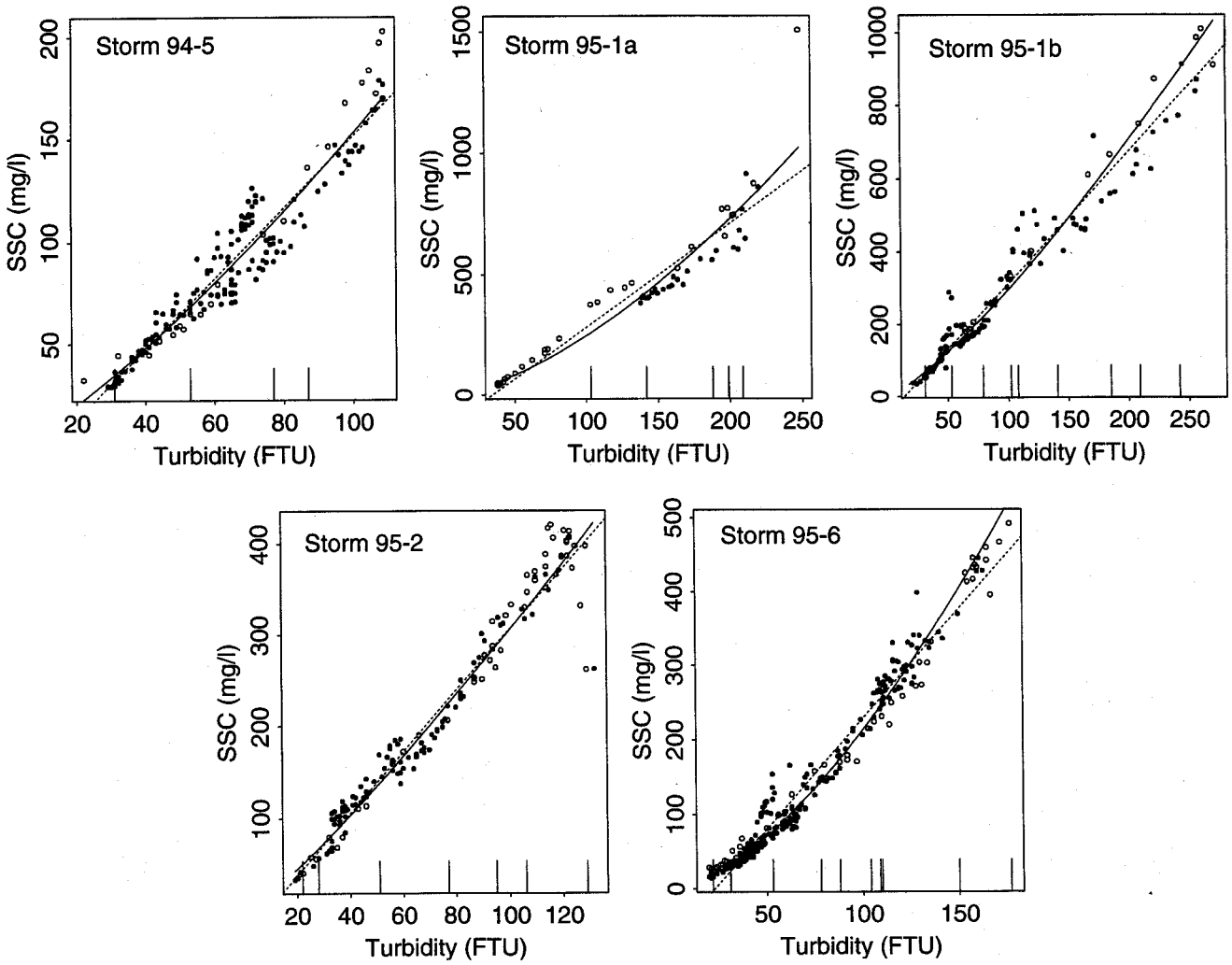

Figure 9. Linear (dotted line) and log linear (solid line) fits of SSC to turbidity for phase 2 storms. The vertical segments along the abscissa indicate turbidity values for which 5000 random SSC values were simulated according to the log linear model. Open circles precede turbidity peak and, solid circles follow it.

continuous records of turbidity using in situ devices requiring little power. By itself, such a turbidity record is of limited utility, because turbidity is very sensitive to variations in the size distribution and composition of suspended solids. Nonetheless, in most streams, variations either are generally not large or are related to SSC, and the relation between turbidity and SSC may be quite stable and precise within bounded time periods. Supplemented with selected concentration specimens, therefore, a continuous turbidity record could provide an efficient method for estimating transported suspended loads. To automate a turbidity-controlled sampling algorithm, a data logger can be programmed to signal a pumping sampler to collect SSC specimens at specific turbidity thresholds for laboratory determination of sediment concentration.

Table 5. Variance Simulation Results

	Sediment Load				Variance					
	% Bias		% rmse		% Bias		% rmse		Standard Error, % rmse	
Storm	Linear	Log	Linear	Log	Linear	Log	Linear	Log	Linear	Log
94-5	0.10	-0.13	6.52	6.43	1.7	3.0	120	110	51	50
95-la	-3.11	-0.14	8.28	7.93	119.0	5.5	244	98	72	43
95-lb	-0.39	-0.13	6.46	6.50	64.2	1.6	150	61	50	29
95-2	0.80	-0.11	5.60	5.21	-16.0	0.6	73	68	38	32
95-6	4.01	-0.02	7.01	5.24	121.4	0.7	199	52	63	25

Variance simulation results comparing linear and log-linear regression (MVUE) estimators as applied to SSC values generated at fixed turbidity levels according to a log linear model with lognormal errors.

Repeated sampling of five extremely dense data sets has demonstrated that runoff event suspended loads in a 384-ha watershed may be estimated consistently to within 8% or better with a turbidity threshold sampling algorithm that results in an average of only 4 to 11 SSC measurements per event. Thresholds that are uniformly spaced after taking square roots provide reasonable sample sizes over a wide range of event magnitudes. Square root scaling resulted in better load estimates than cube root or logarithmic scaling, probably because of its greater emphasis on high turbidities. Because an average of 75% of the sediment is delivered after the turbidity peak, a denser set of thresholds is applied when turbidity is falling. A simple linear regression of SSC on turbidity for each event is usually adequate for accurate estimation of the load. At Caspar Creek the loads are estimated by summing the product of discharge and predicted SSC at 10-min intervals. When the sample data clearly suggest it, load estimates can be further improved with curvilinear fits or individual fits for rising and falling turbidities. However; caution should be exercised in applying nonlinear fits or multiple fits, particularly in the presence of outliers. Extrapolation of nonlinear curves can lead to large errors. And dividing the data is inefficient unless there really are multiple relations.

Although these simulations demonstrate the efficiency of turbidity-controlled sampling in a small stream with predominantly fine sediment transport, the results should be applied with caution in different environments. Ideally, a period of intensive turbidity and SSC measurement followed by this sort of investigation would precede monitoring programs at other sites. At minimum, turbidity records need to be examined in order to determine appropriate thresholds and to verify algorithms for threshold and reversal detection. In the absence of a pilot study with detailed records of both turbidity and SSC, an indication of error magnitude is still available from the variance estimates which can be computed from the operational data. But the uncertainty in the variance estimates is large for small sample sizes, and the variance estimates can be severely biased if the error assumptions of the regression model are not satisfied, or if the relation is not truly of the specified form.

While this paper has focused on estimating suspended sediment loads, the method described has other potential applications. For example, specific conductance could be used in place of turbidity to control sampling for solute load estimation. The approach should be effective for any water quality constituent whose concentration is better correlated with an easily measured (in situ) parameter, such as turbidity or conductance, than with discharge.

Acknowledgments. Phase 1 of this study was conducted under the direction of Robert B. Thomas. Rand Eads was responsible for the field installation. Without their collaboration this research would not have been carried out. I would also like to thank Elizabeth Keppeler for managing the field data collection, David Thornton for managing the sediment laboratory, and the numerous individuals who worked many odd hours collecting and processing SSC specimens.

References

American Public Health Association (APHA), Standard Methods for the Examination of Water and Wastewater, 16th ed., 1269 pp., Am. Public Health Assoc., Washington, D. C., 1985.

Bailey, E. H., W. P. Irwin, and D. L. Jones, Franciscan and related rocks, and their significance in the geology of Western California, *Calif. Div. Mines Geol. Bull*., 183, 177 pp., 1964.

Bogen, J., Monitoring grain size of suspended sediments in rivers, in *Erosion and Sediment Transport Monitoring Programmes in River Basins*, edited by J. Bogen, D. E. Walling, and T. J. Day, *IAHS Publ. 210*, pp. 183-190, 1992.

Cohn, T. A., Recent advances in statistical methods for the estimation of sediment and nutrient transport in rivers, *U.S. Natl. Rep. Int. Union Geod. Geophys. 1991-1994*, Rev. Geophys., *33*, 1117-1123, 1995.

Cohn, T. A., L. L. DeLong, E. J. Gilroy, R. M. Hirsch, and D. K. Wells, Estimating constituent loads, *Water Resour. Res.*, *25*(5), 937-942, 1989.

Colby, B. R., and C. H. Hembree, Computations of total sediment discharge, Niobrara River, near Cody, Nebraska, *U.S. Geol. Surv. Water Supply Pap*. 1357, 1955.

Duan, N., Smearing estimate: A nonparametric retransformation method, J. *Am. Star. Assoc*., 78(383), 605-610, 1983.

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

Fleming, G., and T. Poodle, Particle size of river sediments, J. *Hydraul. Div. Am. Soc. Civ. Eng.*, *96*, 431-439, 1970.

Foster, I. D. L., R. Millington, and R. G. Grew, The impact of particle size controls on stream turbidity measurements; Some implications for suspended sediment yield estimation, in *Erosion and Sediment Transport Monitoring Programmes in River Basins*, edited by J. Bogen, D. E. Walling, and T. J. Day, *IAHS Publ. 210*, pp. 51-62, 1992.

Frostick, L. E., I. Reid, and J. T. Layman, Changing size distribution of suspended sediment in arid-zone flash floods, *Spec. Publ. Int. Assoc. Sedimentol*., 6, 97-106, 1983.

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

Gilvear, D. J., and G. E. Petts, Turbidity and suspended solids variations downstream of a regulating reservoir, Earth Surf. Processes Landforms, *10*, 363-373, 1985.

Gippel, C. J., Dissolved organic stream water colouration in a small forested catchment near Eden, NSW: Source and characteristics, *Working Pap. 1987/3*, 46 pp., Dep. of Geogr. and Oceanogr., Univ. Coll., Aust. Def. Force Acad., Canberra, 1987.

Gippel, C. J., The use of turbidimeters in suspended sediment research, *Hydrobiologia*, *176/177*, 465-480, 1989.

Gippel, C. J., Potential of turbidity monitoring for measuring the transport of suspended solids in streams, *Hydrol. Processes*, 9, 83-97, 1995.

Hadley, R. F., R. Lal, C. A. Onstad, D. E. Walling, and A. Yair, Recent developments in erosion and sediment yield studies, Tech. Dev. Hydrol., Working Group ICCE IHP-II project A.1.3.1, 127 pp., United Nations Educ., Sci., and Cult. Org., Paris, 1985.

Hasholt, B., Monitoring sediment load from erosion events, in *Erosion and Sediment Transport Monitoring Programmes in River Basins*, edited by J. Bogen, D. E. Walling, and T. J. Day, *IAHS Publ. 210*, pp. 201-208, 1992.

Jansson, M. B., Turbidimeter measurements in a tropical river, Costa Rica, in *Erosion and Sediment Transport Monitoring Programmes in River Basins*, edited by J. Bogen, D. E. Walling, and T. J. Day, *IAHS Publ. 210*, pp. 71-78, 1992.

Johnson, A. G., and J. T. Kelley, Temporal, spatial, and textural variation in the mineralogy of Mississippi River suspended sediment, *J. Sediment. Petrol*., *54*, 67-72, 1984.

Krause, G., and K. Ohm, A method to measure suspended load transports in estuaries, *Estuarine Coastal Shelf* Sci., *19*, 611-618, 1984.

Lawler, D. M., M. Dolan, H. Tomasson, and S. Zophoniasson, Temporal variability of suspended sediment flux from a subarctic glacial river, southern Iceland, in *Erosion and Sediment Transport Monitoring Programmes in River Basins*, edited by J. Bogen, D. E. Walling, and T. J. Day, *IAHS Publ. 210*, pp. 233-243, 1992.

Long, Y., and N. Qian, Erosion and transportation of sediment in the Yellow River basin, *Int.* J. *Sediment. Res.*, *1*, 2-38, 1986.

Ongley, E. D., M. C. Bynoe, and J. B. Percival, Physical and geochemical characteristics of suspended solids, Wilton Creek, Ontario, *Can.* J. *Earth* Sci., *18*, 1365-1379, 1982.

Peart, M. R., and D. E. Walling, Particle size characteristics of fluvial suspended sediment, in *Erosion and Sediment Transport Monitoring*

Programmes in River Basins, edited by J. Bogen, D. E. Walling, and T. J. Day, *IAHS Publ 210*, pp. 397-407, 1992.

Reid, L, and L. E. Frostick, Flow dynamics and suspended sediment dynamics in arid zone flash floods, *Hydrol Process., 1*, 239-253, 1987.

Richards, K., Some observations on suspended sediment dynamics in Storbregrova, Jotunheimen, *Earth Surf. Processes Landforms*, 9, 101-112, 1984.

Seber, G. A. F., *Linear Regression Analysis*, 465 pp., John Wiley, New York, 1977.

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

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

Thomas, R. B., and J. Lewis, An evaluation of flow-stratified sampling for estimating suspended sediment loads, J. *Hydrol., 170*, 27-45, 1995.

Wall, G. J., and L. P. Wilding, Mineralogy and related parameters of fluvial suspended sediments in northwestern Ohio, J. *Environ. Qual.*, 5, 168-173, 1976.

Walling, D. E., and P. Kane, Temporal variations in suspended sediment properties, in *Recent Developments in the Explanation and Prediction of Erosion and Sediment Yield*, edited by D. E. Walling, *IAHS Publ. 137*, pp. 409-419, 1982.

Walling, D. E., and P. W. Moorehead, Spatial and temporal variation of the particle-size characteristics of fluvial suspended sediment, *Geogr. Ann., 69A*, 47-59, 1987.

Weaver, C. E., Variability of a clay river suite, J. *Sediment. Petrol., 37*, 971-974, 1967.

J. Lewis, Redwood Sciences Laboratory, 1700 Bayview Dr., Arcata, CA 95521.

(Received December 5, 1995; revised March 15, 1996;
accepted March 28, 1996.)