WATER RESOURCES RESEARCH, VOL. 29, NO. 3, PAGES 583–597, MARCH 1993

A New Model for Bed Load Sampler Calibration to Replace the Probability-Matching Method

ROBERT B. THOMAS AND JACK LEWIS

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

In 1977 extensive data were collected to calibrate six Helley-Smith bed load samplers with four sediment particle sizes in a flume at the St. Anthony Falls Hydraulic Laboratory at the University of Minnesota. Because sampler data cannot be collected at the same time and place as "true" trap measurements, the "probability-matching method" was used to derive surrogate pairs for calibration analysis. The method is invalid since it implicitly assumes that sampler and trap data have no sampling or measurement errors, it gives biased and highly variable results, and it does not contain information enabling model specification. A new calibration model was developed that regresses transformed individual sampler measurements on daily means of transformed trap data and incorporates within-day variation in trap rates to explain part of the sampler variation. Three small-nozzle samplers performed more uniformly than three large-nozzle samplers did. There is evidence that samplers with higher nozzle ratios collect more bed load in most particle size classes tested. However, between the two small-nozzle samplers with ratios of 3.22 and 1.40, significant differences could be detected for only one particle size. The standard sampler with a 76 × 76 mm nozzle trapped sediment less efficiently than a similar sampler with a 152 × 152 mm nozzle in three of four particle sizes tested. Limitations in the data restricted more definitive statements about the samplers, but the results of this study can be used to design a more rigorous calibration experiment.

INTRODUCTION

Bed load samplers are usually calibrated by comparing their "catches" to measurements of "true" bed load transport falling through openings ("traps") installed in a streambed or flume. Calibration is complicated because bed load flux is episodic even under constant hydraulic conditions, often with bed material formed into dunes [*Gomez et al.*, 1989]. This process produces sets of sampler and trap measurements having marginal distributions with large positive skew ranging from zero to a level depending on the hydraulic characteristics of channel and sampler.

Calibration is also hampered by the impossibility of making matched individual sampler and trap measurements of sediment transport at the same time and place on a streambed or flume. Investigators have therefore matched sampler and trap averages made over time under stable hydraulic conditions [*Emmett*, 1980]. *De Vries* [1973], however, pointed out that if the calibration relationship is not linear, it cannot be inferred "without error" using averages.

These problems prompted a flume study at the St. Anthony Falls Hydraulic Laboratory over 10 years ago [*Hubbell et al.*, 1987] to calibrate two standard and four modified Helley-Smith samplers [*Helley and Smith*, 1971]. Instead of matching trap and sampler means a method was developed to derive surrogate pairs of individual measurements. We show why this technique is invalid and develop a new model to calibrate bed load samplers for this experimental protocol.

ST. ANTHONY FALLS FLUME STUDY

Physical Setup and Operation of the 1977 Experiment

Measurements were made in a 2.74 m wide, 1.83 m deep, and 82.91 m long horizontal flume. Water was diverted from the Mississippi River and returned to the river after passing through the flume. Water discharge was controlled at several levels ("runs") using a sluice gate at the head of the flume and a weir at its downstream end.

Sieved bed material was recaptured after measurement at a station 17.98 m upstream from the weir. It was moved by an auger under the measurement devices to a sump at the side of the flume, and pumped through a pipe to the head of the flume and returned to the flow. Thus measurements could be made over long periods under many hydraulic conditions without requiring excessive amounts of sediment. The moving layer of sediment was measured as it fell through seven rectangular traps arrayed across the full width of the flume floor. Sediment passing through each trap fell into a submerged metal "weigh pan" connected by struts passing upward through the flow to a load cell mounted above the flume. Material accumulating in the weigh pans was measured over time by changes in voltage output of the load cells. The bottom of each weigh pan had two hydraulically actuated doors that automatically opened when the pan was full, emptying sediment onto the auger below.

Four particle size distributions of bed material were used. Three nearly uniform-sized materials of 2.1, 6.5, and 23.5 mm were prepared by sieving. The fourth material, a mixture of the other three, formed a nearly lognormal distribution with particles ranging from 1.4 to 32 mm.

Four to six runs (each lasting from 4 to 9 days) were made under nominally constant hydraulic conditions for each particle size distribution. The flume was run for a period after each change in hydraulic or sedimentologic conditions to attain a steady rate of sediment transport. A stationary layer of sediment about 0.46 m thick formed in the flume just upstream of the traps.

Measurements were taken with six designs of Helley-Smith sampler (Table 1) from a movable platform spanning the flume and positioned from 3.0 to 8.5 m upstream from the traps depending on the run. From one to six sampler types were used for a given run; they were operated in rotation so

Paper number 92WR02300.

TABLE 1. Physical Characteristics of Helley-Smith Bed Load Samplers Calibrated and Compared in This Analysis

Sampler	Intake Nozzle Size (Width Times Height in Millimeters)	Nozzle Exit to Entrance Area Ratio	Hydraulic Efficiency*
1	76 × 76	3.22	1.54
2	76 × 76	1.10	1.15†
3	76 × 76	1.40	1.35†
4	305 × 152	1.10	1.15†
5	305 × 152	1.40	1.40†
6	152 × 152	3.22	1.54

Data are from *Hubbell et al.* [1987].
*Ratio of mean flow velocity through the sampler to that which would have occurred through the same cross section without the sampler.
†Estimated.

that repeat measurements for each sampler were distributed fairly uniformly throughout each measurement day.

Measurements were taken of water stage, water surface profile, bed material profile, particle size, and, for most runs, bed material elevation at the sampling station and 1.52 m upstream from the trap. The experiment is described and information on access to data is given in the work by *Hubbell et al.* [1987].

Data

Trap data were collected throughout the daily time periods that the flume was considered to be operating at equilibrium. The weights of accumulated sediment measured by the load cells were recorded continuously on charts and at 6-s intervals on magnetic tape. Differences in the 6-s pan weights from magnetic tape were summed and converted to transport rates for periods of the same length as the Helley-Smith measurement times for that run (18–300 s). Transport rates for trap and sampler were not calculated for the same actual periods of time, but for periods of the same length in the same run.

The trap data were essentially contiguous, while Helley-Smith sampling was interrupted by raising and lowering the devices, record keeping, and emptying and rotating samplers among those used in that run. Because of this the trap data sets were more than an order of magnitude larger than the corresponding data sets for each of the six hand samplers. The trap data therefore essentially represent complete daily populations (except for a few missing values) with known means and variances.

The samplers were usually operated in the center of the flume in line with pan 4. During several runs for the 6.5-mm particle size material the samplers were also used in line with traps 2 and 6.

PROBABILITY-MATCHING METHOD

Individual Helley-Smith sampler values could not be matched in time or space with trap data, so the earlier analysts derived surrogate pairs using the probability-matching method (PMM) [*Hubbell et al.*, 1981; *Hubbell and Stevens*, 1986] hinted at by *de Vries* [1973]. The PMM consists of forming marginal empirical distribution functions from the different-sized trap and sampler data sets and making inverse estimates of sediment transport rates at 100 evenly spaced probabilities from 0.01 to 1. (The usual step-function form of empirical distribution function could not be used because it does not have unique inverses as required by the PMM [*Gibbons*, 1971]. Instead, empirical distributions were formed by connecting adjacent points defined by probabilities at multiples of $1/n$ and their corresponding transport rates with straight line segments.) Trap and sampler transport rates for corresponding empirical probabilities were then paired and plotted. Since the resulting curve is always monotonically increasing, it defines a unique trap rate for every value within the range of the sampler data set. No curve-fitting or smoothing algorithms were applied to these often jagged PMM curves.

Consequences of Using the Probability-Matching Method

General effects of applying the PMM are illustrated by plotting computer-generated data pairs. Fifty bivariate pairs with marginal distributions $X \sim N(3.8, 1.21)$ and $Y \sim N(4.1, 0.81)$ and correlation -0.7 were generated (Figure 1*a*). The results of a simplified PMM can be seen by plotting pairs of identically ranked X and Y observations (Figure 1*b*). (In contrast, the general PMM matches percentiles instead of observations, because the marginal sample sizes are usually unequal.) Another set of Y values from the same marginal distribution was selected for the same set of X values but with correlation 0.0. (Figure 1*c*), and the PMM applied again (Figure 1*d*).

The PMM changed negatively correlated and uncorrelated data into sets of pairs with positive correlations of higher magnitudes and lower variances. The models were distorted and information available in the data that would enable determining their forms and parameters was lost.

The effects of the PMM on positively correlated data are similar if less dramatic. The set of power functions

$$Y = aX^b$$

with positive a and b comprises a rich set of possible linear and curvilinear calibration models passing through the origin. Logarithmic transformation of this set of functions forms a subset of linear functions with positive slopes. Assuming that there are errors in both $\log X$ and $\log Y$ the linear functions can be modeled as a bivariate normal distribution. If the marginal distributions are $\log X \sim N(\mu_X, \sigma_X^2)$ and $\log Y \sim N(\mu_Y, \sigma_Y^2)$ by standard multivariate theory the calibration in log space is

$$E[\log Y|\log X] = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (\log X - \mu_X)$$

in which ρ is the correlation coefficient [*Hogg and Craig*, 1970]. The marginal means and variances define a subset of models, but for given marginal parameters, $0 < \rho \leq 1$ describes an infinite set of possible calibration curves, each in one-to-one correspondence with a unique power function.

The marginal means and variances can be estimated with unpaired data, but marginal distributions do not contain information needed to estimate the correlation. The cross product is required to estimate ρ and real pairs are needed to calculate the cross product. Using the PMM to form n surrogate data pairs selects from the $n!$ possible pairings the

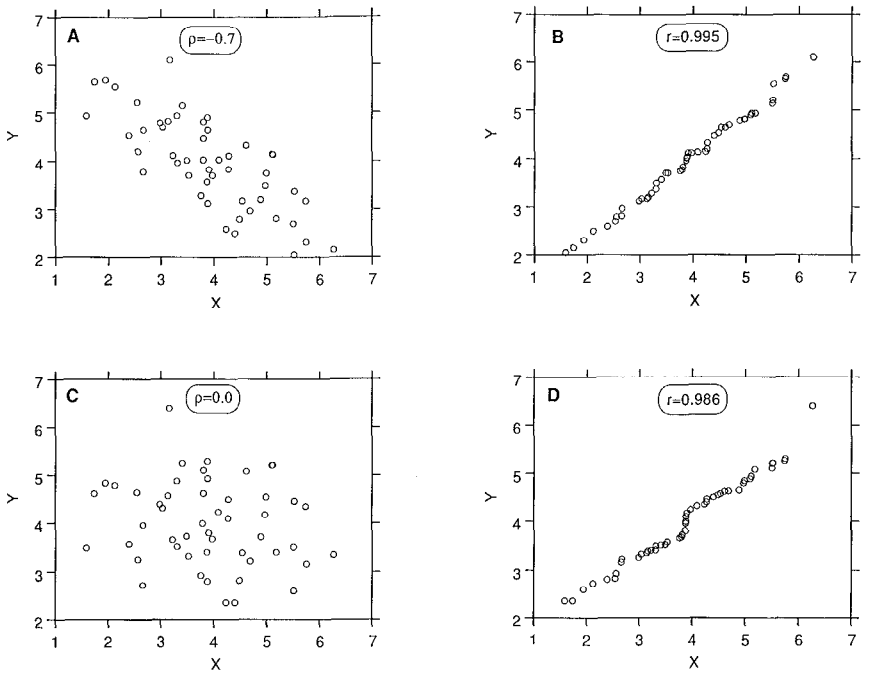

Fig. 1. Computer-generated sets of data pairs with marginal normal distributions $X \sim N(3.8, 1.21)$ and $Y \sim N(4.1, 0.81)$. (*a*) Set of 50 pairs with correlation −0.7. (*b*) Results from transforming the data using the probability matching method. (*c*) Another 50 pairs using the same X values as in Figure 1*a* but with correlation of 0.0. (*d*) Results from applying the probability matching method to the uncorrelated data.

one that maximizes the cross product and therefore the correlation estimate for that data (inductive proof available from authors).

The effects of the PMM on regression are illustrated by a set of 50 bivariate normal pairs generated with correlation 0.6 and the same marginal distributions described earlier (Figure 2). The untransformed data are plotted as open circles and their least squares regression is shown by the solid line. The PMM-transformed data are shown by pluses and their regression by the dashed line. This estimated regression line is steeper and the variance of the points about the line is smaller as indicated by the narrower pattern of pluses.

This result is seen to be general by expressing the estimated regression slope $\hat{\beta}$ as

$$\hat{\beta} = r\hat{\sigma}_Y/\hat{\sigma}_X$$

in which the estimated standard deviations $\hat{\sigma}_Y$ and $\hat{\sigma}_X$ are not affected by the PMM [*Weisberg*, 1985]. Because it maximizes the correlation estimate r, the PMM produces a set of points that causes the regression line to rotate counterclockwise around the means. The residual sum of squares (RSS) is minimized since

$$RSS = SYY(1 - r^2)$$

in which SYY is the sum of squared deviations of the Y values from their mean and is also not affected by the PMM [*Weisberg*, 1985]. Because the estimate of the regression variance is proportional to the RSS, it, too, is minimized. Therefore using the PMM on a set of bivariate pairs maximizes the slope and correlation estimates and minimizes the variance estimates for that set of data. The intercept is maximized or minimized according to whether the sample mean of X is negative or positive respectively.

The PMM restores correct pairing from a set of "severed" pairs only when a data set is "monotonic," that is, when every X value is associated with a Y value at least as large as the Y values paired with all smaller X values. Monotonic

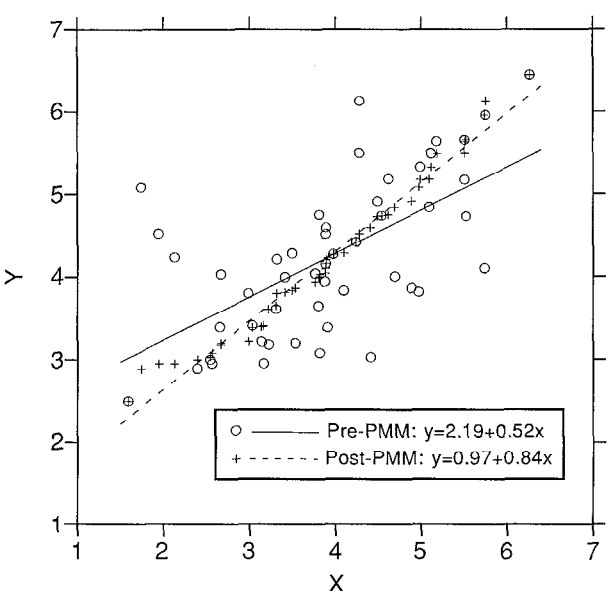

Fig. 2. Scatterplot of 50 data pairs with marginal normal distributions $X \sim N(3.8, 1.21)$ and $Y \sim N(4.1, 0.81)$ and correlation 0.6 are shown by open circles. The estimated least squares regression is shown by the solid line. The pluses show the pairs after transformation by the probability-matching method and the dashed line shows the regression on those data. The probability-matching method increases the slope and correlation of the model and reduces the variance.

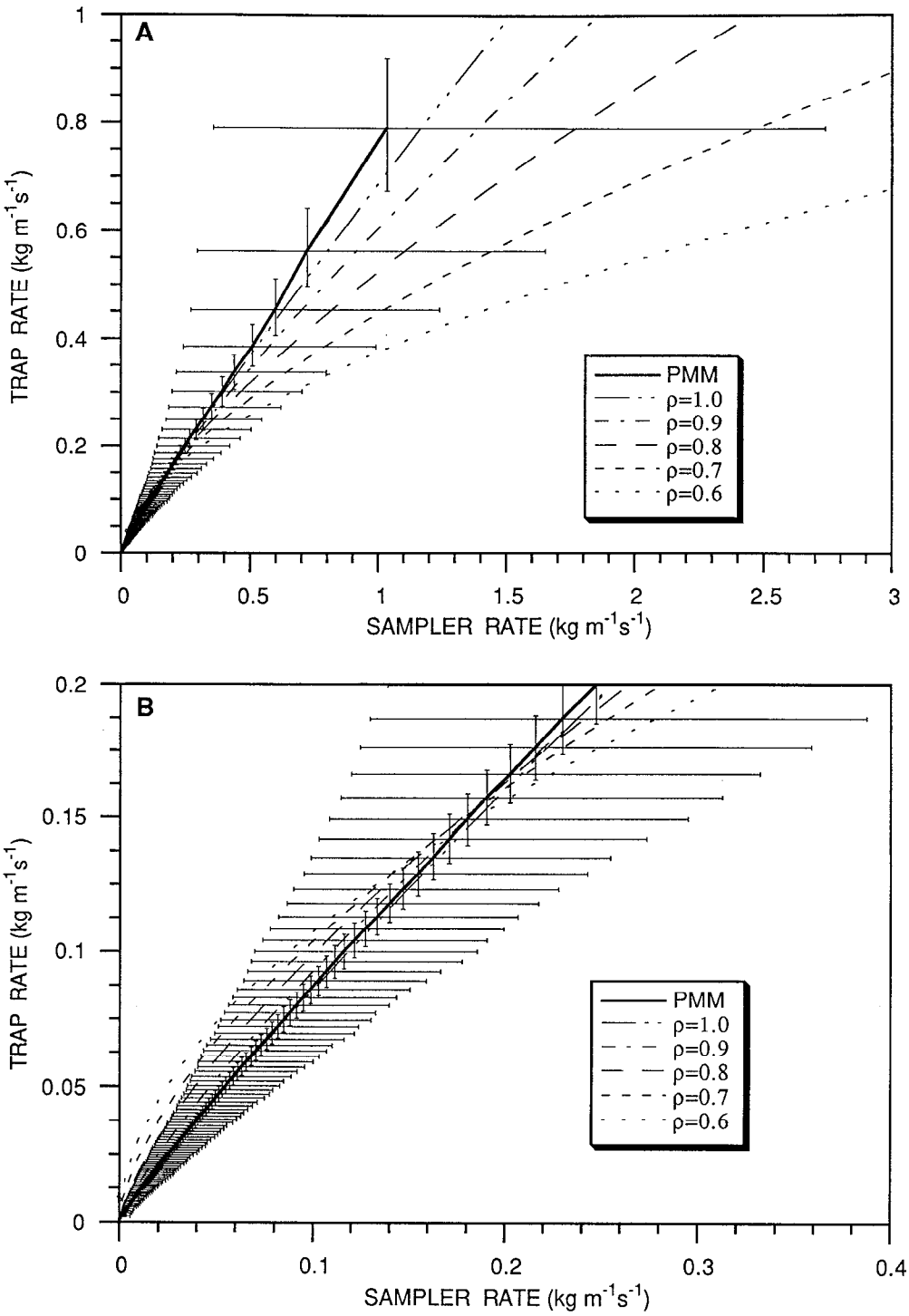

Fig. 3. Simulation of the 99 percentiles of X and Y estimated by the probability-matching method from 1000 samples of 3000 Y and 60 X values sampled from the marginal distributions of a bivariate lognormal distribution. The pairs of means of the corresponding percentiles were connected by straight-line segments. The 95% prediction intervals were determined from the percentile estimates of rank 25 and 975 and were plotted as horizontal and vertical bars around their mean percentiles. Four possible models with correlations of 0.7, 0.8, 0.9, and 1.0 are shown. Figure 3*b* shows only the lower 88 percentiles.

data sets rarely occur; they are virtually always due to chance and are promoted by low variance and widely spaced X values. The PMM can be used without distortion only if bed load calibration data are monotonic.

The PMM produces a quantile-quantile plot. Such plots can be used to study relationships between distributions. However, they cannot be used to develop functions to predict one variate from another for specific values [*Chambers et al.*, 1983]. The information contained in the cross product is essential for this purpose; it is information that the PMM (or any other technique) cannot recover for unpaired data.

Conditions for the Probability-Matching Method

To avoid these problems *Hubbell et al.* [1985] assumed three validating conditions which we restate here. The variables q_t, q_s, and q_p denote the true, sampler, and pan transport rates, respectively.

First, the number n of sampled rates and rates measured at the weigh pans is sufficiently large so that the cumulative relative-frequency distribution, $F_n(q_s)$ or $F_n(q_p)$, of q_s and q_p values, respectively, are equivalent to the probability distribution functions $F_X(q)$ of their populations. That is,

$$F_X(q) = \lim_{n \to \infty} F_n(q) = F_n(q_s) \text{ or } F_n(q_p)$$

in which $F_X(q)$ is the probability that either q_s or q_p is less than or equal to any real rate q.

Second, for any given flow condition, true rates at all longitudinal locations along the flume and rates at the weight pans are identically distributed. That is,

$$P[q_t \leq q] = P[q_p \leq q]$$

or, equivalently,

$$F_{q_t}(q) = F_{q_p}(q)$$

in which P denotes probability.

Third, the relative position (order) of each sampled rate in its distribution is the same as the relative position of the corresponding true rate in its distribution. That is, for every sampled rate q_1 there is a corresponding true rate q_2 that would have occurred at the time and place of sampling had the sampler not been there, such that

$$P[q_s \leq q_1] = P[q_t \leq q_2]$$

The mathematical statement of the first condition requires that the piecewise linear empirical distribution functions (of trap or sampler data) be equal to their corresponding population distributions. (The descriptive assertion requiring "equivalence" has no clear operational meaning so is not discussed.) The first condition effectively prohibits sampling error in either trap or sampler transport rates. But sampling variation is present in all sampling procedures; a major task of calibration is to estimate its magnitude and effects on the estimates.

The intended meaning of the first condition is probably that the sample sizes are large enough to achieve approximate equality. While the condition of equality is essentially impossible to meet, we shall see (Figure 3) that even the less stringent requirement of approximate equality, although vaguely defined, is most certainly not satisfied with the Helley-Smith sample sizes in the St. Anthony Falls data.

The second condition equates the distributions of bed load transport rates at the pans and true rates at any of the Helley-Smith sampling stations for the same run. Although true rates at the sampling stations are not known, the presence of negative trap rates (see the section on transformations) proves they are different. Under steady state conditions, the means of these distributions are likely to be equal, but conditions at the traps suggest that their shapes are different.

The presence of traps alters the bed material profile immediately upstream. Downstream sediment support removed by material falling into the traps forms an "avalanche face" exposing the nonmoving stratum under the dunes. The shape of the dune approaching the traps is also altered by loss of support downstream, changing hydraulic and sedimentologic factors controlling sediment movement into the traps. If pan measurements are a function of bed elevation (as *Hubbell et al.* [1985] suggest in their discussion of the third condition), changing the dune shape should alter the distribution of pan transport rates. Sediment moving off the dunes was observed to accumulate on the avalanche face and then to slough off into the pans at irregular intervals (J. V. Skinner, personal communication, 1991). Such deposits are unlikely to have the same distribution as bed load flux through a vertical cross section in the absence of traps. Measurement errors may also result from pressure waves due to falling sediment, hydraulic eddies, vertical flow components, pan emptying, and the load-cell transducer system. These processes differ from those at the sampling site so it would be remarkable if the distributions were identical.

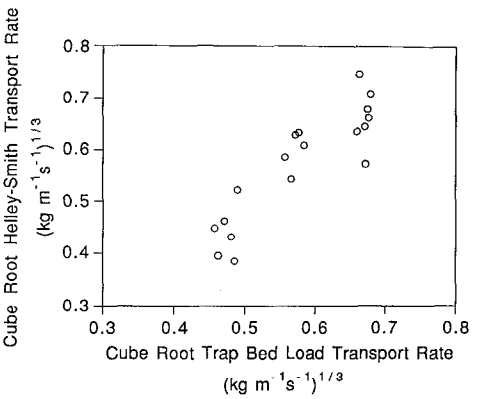

Fig. 4. Typical transformed bivariate plot of Helley-Smith and trap means illustrating linearity (showing sampler five with the mixed particle size).

Since the second condition requires the distribution of pan rates to be identical to the distribution of true rates everywhere in the flume, the condition effectively requires that there be no measurement errors in sediment transport rates determined at the traps. Errors are present in all physical measurements. To find a situation with no measurement errors (especially in a process as difficult to measure as bed load transport) would be especially startling. A basic purpose of calibration is to use observed variation of matched pairs to specify calibration model form and to estimate prediction error.

The third condition is the crux of the PMM, requiring that there be no measurement errors in the sampler rates. While the true and sampler rates do not have to equal each other, a given true rate must always be matched with a particular sampler rate. Therefore the Helley-Smith sampler must behave identically under any combination of local hydraulic and sedimentologic conditions that produce a given true transport rate. It disallows obvious sources of measurement error from such factors as positioning and maintaining the sampler on the dune, timing errors, and errors in emptying the sampler bag.

None of the three "PMM validating" conditions is acceptable. Sampling and measurement errors occur in all data and

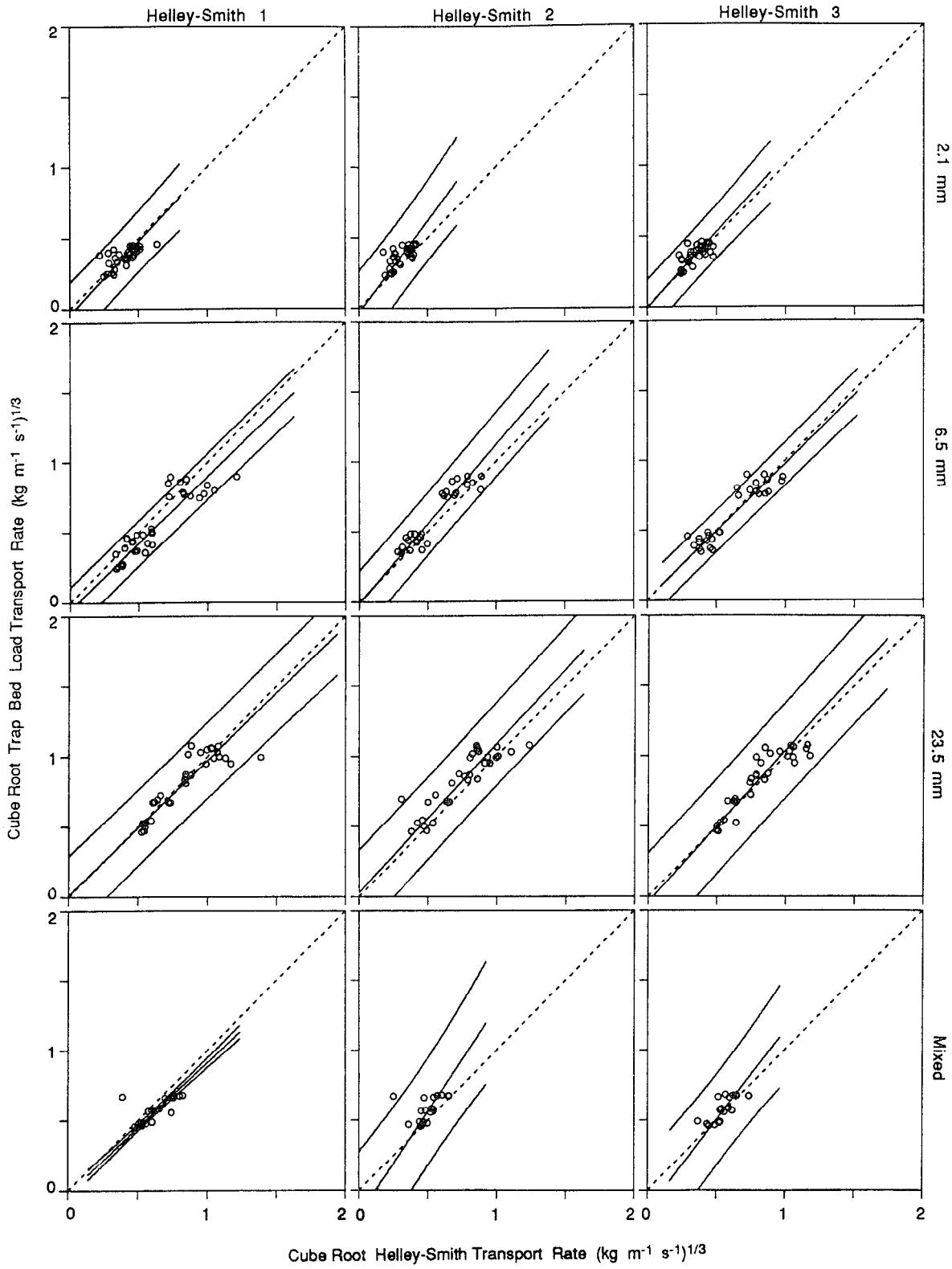

Fig. 5. Transformed inverse Helley-Smith calibration lines, discrimination intervals, and dashed lines of equal response ($y = x$). Lines extend to the extremes of the individual Helley-Smith data used in each calibration. Circles are daily means of the transformed Helley-Smith and trap data.

difficulty in pairing does not justify assuming that they do not exist. The PMM implicitly denies these universal properties of empirical measurement and therefore should not be used with any data.

Simulation

Finally, computer simulation was used to investigate the performance of PMM under sampling variation. Sampler 6

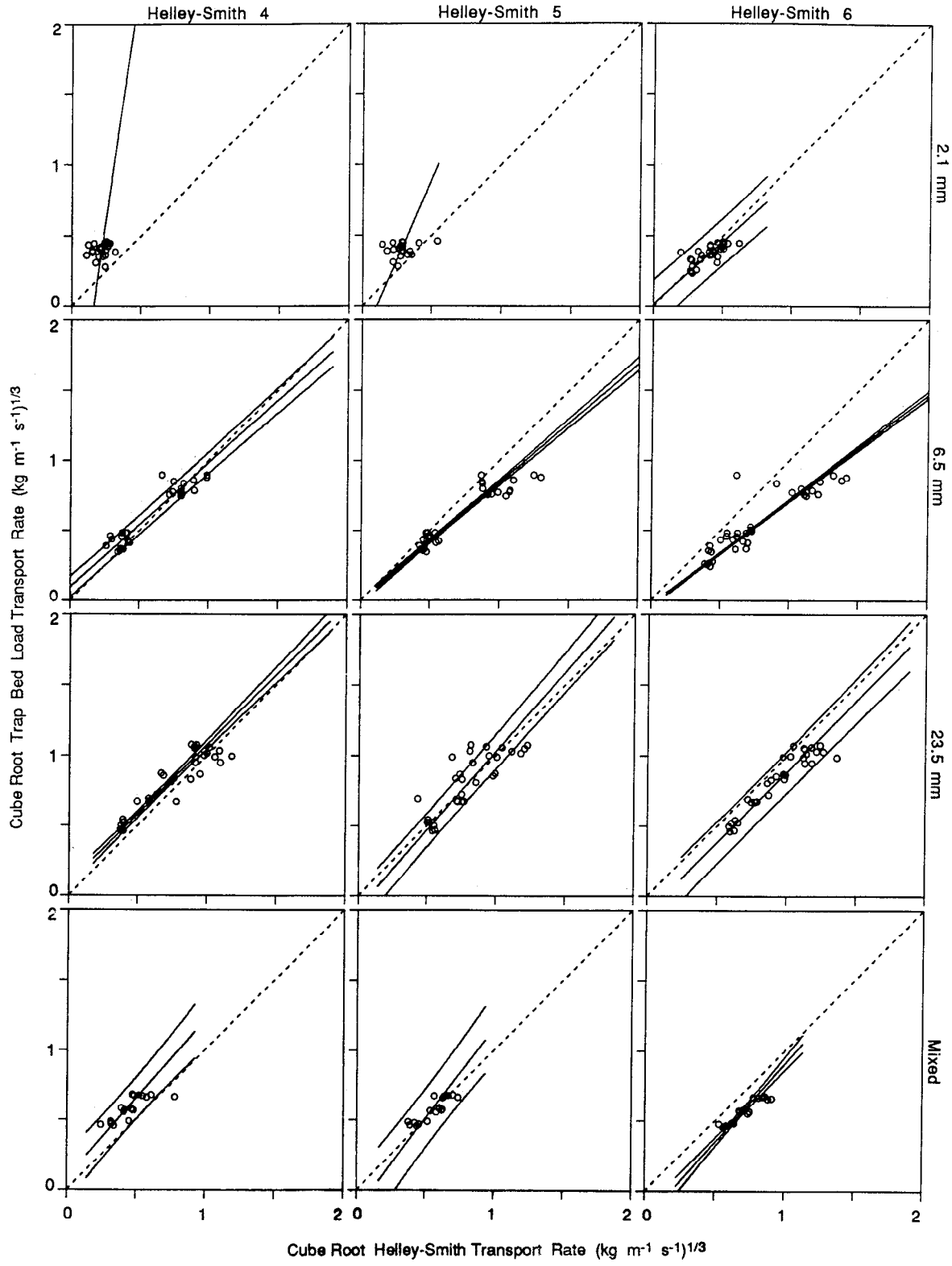

Fig. 5. (continued)

from run 4 with the 6.5-mm material was selected because the distributions of both sampler and pan data were approximately lognormally distributed and typical of the data collected. The sample mean of the log-transformed trap data was −3.146 with sample variance 1.568 and corresponding values for the sampler data were −3.065 and 1.918 respectively. We generated 3000 trap and 60 "sampler" values for each simulation and estimated the 99 percentiles of the marginal distributions by interpolation between the ordered sample values (the hundredth percentiles of the marginal distributions are infinite). The simulation was repeated 1000 times and the percentile estimates of rank 25 and 975 were

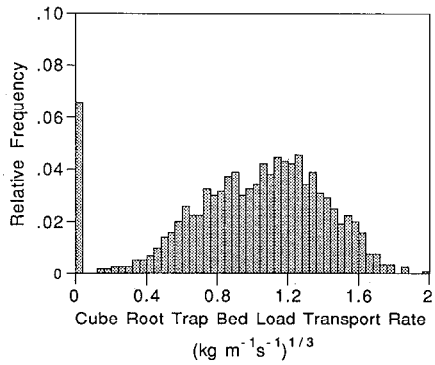

Fig. 6. Typical histogram of cube root transformed daily trap transport rates showing stack of points at zero. Particle size is 23.5 mm.

used to estimate 95% confidence intervals for each trap and sampler percentile, which are plotted as vertical and horizontal bars around the means of the corresponding percentile pairs (Figure 3). By varying the correlation coefficient we show five models that could have produced the sampled marginal distributions.

Mean PMM coordinates for the top 15 percentiles lie above all of the underlying models (Figure 3*a*) and, for the lower 83 percentiles, lie below all models except the one with correlation 1.0 (Figure 3*b*). Because the PMM does not respond to different values of correlation that describe the true model, it returns the same set of estimates for each of the five models. Most striking, however, is the high variability in the placement of the mean pairs as indicated by the lengths of the confidence bars. Sample sizes much larger than 60 are needed to precisely estimate percentiles, particularly in the upper tail. Not only is the PMM insensitive to the underlying model, but it produces highly erratic results with practicable sample sizes.

NEW MODEL

Because the PMM is problematic, other approaches to analysis were attempted. Efforts to match the data statistically using lagged correlations and time series analysis were unsuccessful. These attempts probably failed due to large changes in dune size and shape during movement, the distance separating the sampling station and the traps (from 3.05 to 8.54 m), and differences in hydraulic conditions at the two stations.

A solution to *de Vries*'s [1973] problem of using means with curvilinear data is to transform the data to linear form before averaging. Linearly related daily sampler means could be regressed on trap means using weighted regression to account for different Helley-Smith sample sizes and the fitted equations adjusted to predict individual measurements. Unfortunately, this model distorts the error from individual measurements and cannot be used. The following sections investigate these problems in detail and develop a new model to fit the necessary bed load data calibration protocol.

Transformations

Daily distributions of trap transport rates contain large proportions of zero values (including negative values that had been coded to zero in the summarized data that were available to us) and daily distributions of both sampler and trap data show pronounced positive skew. Both of these factors were considered in choosing a transformation to approximate the assumptions of linearity and normality (see the calibration model section).

Linear pairs of mean values do not, in themselves, imply that the pairs of individual measurements comprising them are linear. With a monotonic set of three or four aligned means, it is possible to devise examples (e.g., sigmoid curves) where the individual measurements do not follow a linear relationship. However, with normality of the marginal distributions, and where enough means are available to produce a "linear cloud" on a scatterplot, realistic counterexamples become difficult to imagine. Most plots of the St. Anthony Falls data could be described as linear clouds, but some are more like a pair of circular clouds which result when only two widely separated sets of hydraulic conditions were used in the flume. We speculate that if a wide and uniform range of conditions had been used as we recommend, these clumpy plots would look much like the others. We therefore assumed that having linear means was adequate evidence that the individual observations would also be linear if they could be paired.

After transformation of the data, the new model we propose requires normal marginal distributions and linearity between Helley-Smith values and the daily mean trap transport values. The data were transformed using the "family" of power functions x^λ in which the value of λ controls the "strength" of the transformation and logarithms are used in place of $\lambda = 0$ [*Box and Cox*, 1964; *Hoaglin et al.*, 1983]. These transformations are widely employed in statistical analysis in general and for bed load sampler calibration in particular [*Emmett*, 1980]. Several values for λ were tried and daily means of transformed sampler and trap observations plotted and judged for linearity. The best general value of λ over all 24 plots of means was one third (Figure 4). (A few plots may appear slightly curvilinear (Figure 5), but in every case this impression depends on one or two outlying points which could easily result from high variances due to small Helley-Smith sample sizes.) The cube root transformation also resulted in reasonably symmetrical marginal distributions overall, except for the "stacks" of zeros to which we turn next (Figure 6).

High proportions of zero rates were not present in most of the sampler data (Table 2). The high rates of zeros for two samplers with the 2.1-mm particle size, especially for Helley-Smith sampler 4, may have resulted from small sample sizes and relatively low stream power. But, even for sampler

TABLE 2. Percentage of Zeros in Trap and Helley-Smith Sampler Transport Rates for Four Particle Sizes

	Particle Size, mm			
Device	2.1	6.5	23.5	Mixture
Trap	7.2	8.6	6.2	7.7
HS 1	0.0	0.3	1.3	0.0
HS 2	3.4	0.6	1.9	0.4
HS 3	1.1	0.0	2.1	0.0
HS 4	13.6	0.8	0.0	0.0
HS 5	6.2	0.0	0.0	0.0
HS 6	0.6	0.0	0.0	0.0

HS, Helley-Smith.

4, almost 40% of the days had no zeros, while the traps had zeros every day.

Although data removal should be approached warily, we determined the nonpositive trap rates likely arose from problems peculiar to their measurement. Our reasoning for eliminating zeros from all trap data sets follows. The few zeros in the sets of sampler values were not removed since there was no similar justification for their removal.

While the highly variable nature of sediment flux suggests that true zero trap rates can occur, they are probably rare, especially under high stream power or when the length of the sampling time is increased [*Gomez et al.*, 1989]. Yet from 2.4 to 10% (median 6.8%) of trap rates on any given day were nonpositive. Stacks of these nonpositive flux values (e.g., Figure 6) seem to result from the contamination of the distribution of true trap rates by some external process. Two mechanisms were identified that could account for this contamination.

One mechanism comes from the additional water supplied to the sump to facilitate pumping sediment to the head of the flume (J. V. Skinner, personal communication, 1991). This water was added to prevent the high-capacity sediment-return pump from removing too much water from the flume resulting in downward velocities around the weighing pans and causing positive bias in their readings. To minimize this bias and the risk of "starving" the pump, a separate water supply was introduced into a constant-level water tank above the sump. Occasional manual adjustments were made to a weir to match the level in this tank to the water level in the main flume. Precise adjustment was difficult, however, and the press of work during data collection resulted in occasional downward or upward vertical velocities around the pans until readjustments could be made.

The vertical velocities would have affected rates of any magnitude and could have been responsible for producing negative transport rates of varying sizes. In the summarized data available to us these negative rates had been set to zero, so the measured zeros could not be distinguished from negative values that had been changed to zero. These negative values could account for the stacks of zero flux.

Another mechanism for zero trap transport rates is the behavior of sediment moving onto the avalanche face just upstream from the traps. Sediment was observed to move onto the face more or less continually, accumulate there, and then to slough off into the pans in larger amounts, but less frequently, than movement onto the face (J. V. Skinner, personal communication, 1991). This could account for some of the legitimate zeros. Again, this mechanism would affect rates over the entire spectrum of flux. Accumulation of sediment on the avalanche face was supported by a runs test showing greater "clumping" of zero transport rates than would be expected in a random sequence. However, the clumps tended to be followed by lower than average pan transport rates, not the higher ones this process would seem to imply.

Effects on the trap observations of measurement errors due to vertical velocity currents and avalanche face accumulation are complex and cannot be determined from these data. The major effect of the errors is to increase the variance estimated from the daily trap observations (identified in the new model as sampling variance), although omitting the zero values will counter that effect to some unknown extent. We believe it is likely that trap measurement errors are relatively small and had modest effects on the results.

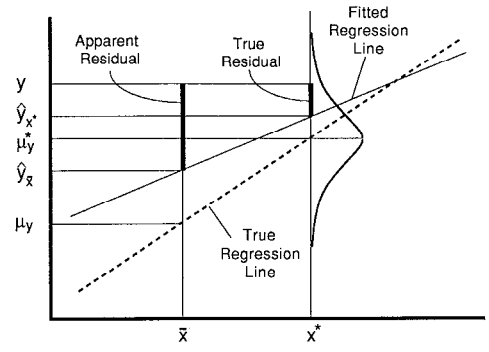

Fig. 7. Illustration of difference between true and apparent residuals (from fitted regression) due to lack of matching individual measurements.

Why Least Squares Regression Is Inappropriate

Since the cube root transformation normalizes and linearizes the data, we considered using least squares regression on the means of the transformed values to obtain calibration lines and prediction bands. The more highly variable means of the transformed Helley-Smith transport rates would be regressed on trap means which can be treated as known without error due to their very large sample sizes. Weighted least squares regression would be used to account for differing numbers of individual rates (and the resulting differing variances) comprising the sampler means. The estimated regression could then be solved for the trap variable and reverse prediction intervals calculated using the standard calibration model [*Miller*, 1981; pp. 117–120].

However, the regression mean squared error cannot be used to estimate the variance needed for prediction bands because the error term in conventional least squares models does not correctly model regression on means. Because individual sampler and trap observations cannot be matched, variation in individual rates comprising the trap means leads to distorted "residuals" in the means model.

To see this, consider the unobserved true transport rate x^* corresponding to a Helley-Smith observation y (Figure 7). Assume this rate is what would have been observed if a trap had been in the flume at the time and location that the Helley-Smith sample was taken. The y values for the rate x^* are assumed to be normally distributed with mean μ_y^* and variance σ^2. The true regression line passes through μ_y^*, while the estimated regression (based on the means) passes through another point $\hat{y}_{x*}$. The true residual from the estimated regression is $y - \hat{y}_{x*}$ (shown by the shorter heavy vertical line segment). Since x^* is unknown, y must instead be matched for analysis with $\bar{x}$, the mean daily trap value. At $\bar{x}$, the true regression line passes through μ_y and the estimated line passes through $\hat{y}_{\bar{x}}$. The apparent residual (shown by the longer heavy vertical line segment) is then $y - \hat{y}_{\bar{x}}$, which differs from the true residual by $|\hat{y}_{x*} - \hat{y}_{\bar{x}}| = \hat{\beta}(x^* - \bar{x})|$, where $\hat{\beta}$ is the estimated regression coefficient. The difference is due to within-day variation in equilibrium trap transport rates. This difference should be included in the explained variance due to regression, rather than the error

TABLE 3. Estimates of Model Parameters $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma}^2$ and Their Variances and Covariance for Making Predictions of Trap Transport Rates and Estimating Inverse Prediction (i.e., Discrimination) Intervals for Each Sampler and Particle Size

Particle Size, mm	Sampler	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\text{V}}\text{ar}[\hat{\alpha}]$	$\hat{\text{V}}\text{ar}[\hat{\beta}]$	$\hat{\text{C}}\text{ov}[\hat{\alpha},\hat{\beta}]$
2.1	HS 1	0.0393	0.965	0.0111	0.00122	0.00990	−0.00337
2.1	HS 2	0.0243	0.770	0.0118	0.00121	0.00974	−0.00334
2.1	HS 3	0.00887	0.930	0.00798	0.00093	0.00777	−0.00261
2.1	HS 4	0.1620	0.144	0.0212	0.00928	0.05694	−0.02284
2.1	HS 5	0.1057	0.443	0.0160	0.00703	0.04385	−0.01742
2.1	HS 6	−0.0147	1.125	0.00892	0.00106	0.00874	−0.00294
6.5	HS 1	0.0562	1.051	0.00745	0.00016	0.00085	−0.00032
6.5	HS 2	0.0102	0.883	0.0102	0.00062	0.00178	−0.00098
6.5	HS 3	−0.00568	1.029	0.00605	0.00054	0.00158	−0.00084
6.5	HS 4	−0.1068	1.124	0.00148	0.00047	0.00139	−0.00073
6.5	HS 5	0.00435	1.180	0.0000*	0.00023	0.00052	−0.00027
6.5	HS 6	0.0550	1.327	0.0000*	0.00006	0.00024	−0.00009
23.5	HS 1	−0.0129	1.042	0.0202	0.00188	0.00321	−0.00230
23.5	HS 2	−0.0260	0.945	0.0197	0.00172	0.00280	−0.00207
23.5	HS 3	0.0383	0.926	0.0251	0.00177	0.00288	−0.00212
23.5	HS 4	−0.0827	1.006	0.0000*	0.00059	0.00092	−0.00064
23.5	HS 5	0.0851	0.890	0.00278	0.00089	0.00157	−0.00110
23.5	HS 6	0.1249	0.991	0.00536	0.00118	0.00196	−0.00141
Mixed	HS 1	0.0210	1.072	0.0000*	0.00059	0.00152	−0.00082
Mixed	HS 2	0.1215	0.674	0.0171	0.00435	0.01352	−0.00755
Mixed	HS 3	0.1122	0.782	0.0165	0.00515	0.01624	−0.00901
Mixed	HS 4	−0.0729	0.885	0.00442	0.00284	0.00876	−0.00488
Mixed	HS 5	0.1129	0.770	0.00673	0.00261	0.00765	−0.00437
Mixed	HS 6	0.1947	0.889	0.0000*	0.00085	0.00255	−0.00138

HS, Helley-Smith.

*These cases maximized the primary likelihood function with negative values of $\hat{\sigma}^2$, so they were recalculated with another likelihood function that found values for $\hat{\alpha}$ and $\hat{\beta}$ while assuming that $\sigma^2 = 0$.

variance. The model we introduce next includes a term for the daily variation in trap rates.

Calibration Model

Because of the large numbers of daily trap observations we assumed that the sample means and variances were the true parameters for normally distributed populations of transformed trap transport rates. A model was developed that matches unobservable trap rates belonging to these populations with individual Helley-Smith rates. This model can be expressed in terms of the known means and variances which eliminates the unobservable quantity. The resulting model is then solved using maximum likelihood to estimate parameters.

For day i of a given run and particle size, let y_{ij} be the jth transformed Helley-Smith observation corresponding to an unobservable transformed trap rate x^*_{ij}. The x^*_{ij} is assumed to be the trap measurement that would have been obtained at the Helley-Smith sampling station if a trap had been there under conditions identical to those existing when the Helley-Smith sample was obtained. Suppose, also, that the Helley-Smith observation is a linear function of x^*_{ij} with measurement error $\varepsilon_{ij} \sim N(0, \sigma^2)$ in the form

$$y_{ij} = \alpha + \beta x^*_{ij} + \varepsilon_{ij} \tag{1}$$

Now let μ_i be the mean trap rate for day i and assume that x^*_{ij} has sampling error $e_{ij} \sim N(0, \delta_i^2)$ independent of ε_{ij} so that

$$x^*_{ij} = \mu_i + e_{ij} \tag{2}$$

Now substitute the expression for x^*_{ij} into (1) to obtain

$$y_{ij} = \alpha + \beta(\mu_i + e_{ij}) + \varepsilon_{ij} \tag{3}$$

Then, because

$$E[y_{ij}] = \alpha + \beta\mu_i \tag{4}$$

$$\text{Var}[y_{ij}] = \beta^2\delta_i^2 + \sigma^2 \tag{5}$$

we have

$$y_{ij} \sim N(\alpha + \beta\mu_i,\ \beta^2\delta_i^2 + \sigma^2) \tag{6}$$

The values of μ_i and δ_i^2 are assumed known and, based on the assumption of normality of the y_{ij}, maximum likelihood [*Hogg and Craig*, 1970] was used to estimate the model parameters α, β, and σ^2. A computer program was written to solve the maximum likelihood equations which used closed-form expressions for the derivatives and the Newton-Raphson method [*Chapra and Canale*, 1988] for optimization.

In five runs of the maximum likelihood program, σ^2 converged to negative values. To ensure that we had not found local maxima due to poor starting values in these cases, we plotted the log likelihood function for maximum likelihood and least squares estimates of α and β. In every case we found a smooth well-behaved function with a single maximum. Since σ^2 cannot be negative, another version of the program was written that maximized the likelihood by choosing optimal values for $\hat{\alpha}$ and $\hat{\beta}$ while assuming $\sigma^2 = 0$. This process introduces some bias, but avoids absurd values of σ^2. Final parameter estimates for the 24 combinations of particle size and Helley-Smith sampler are shown in Table 3.

The fitted line for each sampler and particle size can be written as

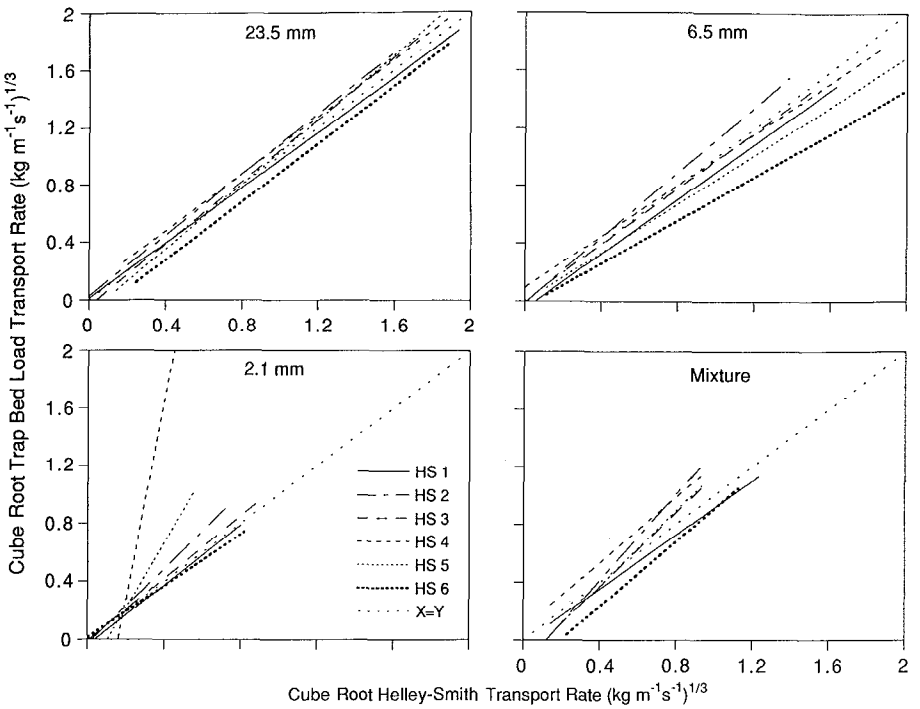

Fig. 8. Transformed calibration lines for all Helley-Smith samplers for each particle size.

$$\hat{y} = \hat{\alpha} + \hat{\beta}x \tag{7}$$

in which the subscripts on x and y are dropped because we are no longer in the flume environment. Then for the cube root y_0 of a Helley-Smith measurement the cube root of the corresponding trap value is estimated by solving (7) for x to obtain

$$\hat{x}_0 = (y_0 - \hat{\alpha})/\hat{\beta} \tag{8}$$

Equation (8) is the calibration relationship to be fit for each of the 24 samplers and particle sizes. The actual estimated trap rate is the cube of $\hat{x}_0$; no correction is needed for transformation bias because individual transport rates (and not means) are being predicted.

Estimated "discrimination intervals" (i.e., inverse prediction intervals) surrounding the calibration curves in (8) can be obtained by simultaneously solving the equations of the confidence bands (for predicting one additional point) around (7) and the line $y = y_0$ [*Miller*, 1981, pp. 117–120]. Approximate intervals can also be computed using the delta method to estimate the variance of a function of random variables [*Bishop et al.*, 1975]. Discrimination intervals based on the delta method are presented since they are easier to compute and except for two pathological cases were almost identical to the "exact" intervals. The method requires estimates of the variances of $\hat{\alpha}$, $\hat{\beta}$, the measured Helley-Smith value y_0, and all covariances which were available from the maximum likelihood process. Because y_0 is a future observation, its covariances with $\hat{\alpha}$ and $\hat{\beta}$ are both zero so that terms containing these factors drop out of the equation. Also, in a prediction situation, $\delta_i^2 = 0$ since we are only interested in the predicted trap rate x_0 at the instantaneous true rate. We do not care about variation in the true transport rates that might exist over some period of time in which hydraulic equilibrium might persist as in the flume. Therefore $\text{Var}[y_0] = \sigma^2$. Then, an estimate of the variance of the predicted trap transport rate is given by

$$\hat{\text{Var}}[\hat{x}_0] = \frac{1}{\hat{\beta}^2}\hat{\text{Var}}[\hat{\alpha}] + \left(\frac{y_0 - \hat{\alpha}}{\hat{\beta}^2}\right)^2 \hat{\text{Var}}[\hat{\beta}] + \frac{\hat{\sigma}^2}{\hat{\beta}^2} + 2\left(\frac{y_0 - \hat{\alpha}}{\hat{\beta}^3}\right)\hat{\text{Cov}}[\hat{\alpha}, \hat{\beta}] \tag{9}$$

The square root of the $\text{Var}[\hat{x}_0]$ multiplied by the cutoff point of the normal distribution for the desired confidence level is added and subtracted from $\hat{x}_0$ in (8) for particular values of y_0. That is, for approximate 95% discrimination intervals, the bands are at

$$\hat{x}_0 \pm 1.96(\hat{\text{Var}}[\hat{x}_0])^{1/2}$$

Estimated values of σ^2 and the variances of $\hat{\alpha}$ and $\hat{\beta}$ and their estimated covariances are given in Table 3. Plots of the calibration lines, their 95% discrimination intervals, the line of equal response ($y = x$), and the mean daily Helley-Smith and trap data are given for all six samplers and four particle sizes in Figure 5. The lines are plotted to the extremes of the individual Helley-Smith data used in each calibration. Plots of all samplers for each particle size give another view for comparison (Figure 8).

Helley-Smith Comparisons

Four sets of formal comparisons were made to assess differences between samplers. The null hypotheses tested were H_0: $\alpha_i = \alpha_j$ and H_0: $\beta_i = \beta_j$ and, for the fourth set of comparisons, H_0: $\alpha_i = 0$ and H_0: $\beta_i = 1$. We assumed for these tests that the maximum likelihood estimators are normally distributed, which is always true asymptotically [*Kendall and Stuart*, 1967]. Since estimates of the parame-

TABLE 4. Statistics for Testing Equality of Calibration Parameters Among Uniform Particle Sizes and Between the Average of the Uniform Particle Sizes and the Mixed Particle Size

Particle Size Contrast	Parameter	Helley-Smith Sampler 1	2	3	4	5	6
2.1/6.5	α	−0.454	0.329	0.379	2.722	1.190	−2.082
	β	−0.834	−1.052	−1.024	−4.058*	−3.497*	−2.137
2.1/23.5	α	0.937	0.928	−0.566	2.462	0.232	−2.949*
	β	−0.677	−1.556	0.043	−3.583*	−2.098	1.293
6.5/23.5	α	1.529	0.748	−0.915	−1.112	−2.413	−1.985
	β	0.139	−1.268	1.784	2.658	6.331*	7.170*
Average/mixed	α	−0.211	1.722	1.325	−1.009	0.806	4.201*
	β	0.959	−1.564	−1.355	1.021	−0.597	−4.207*

*Significant at the 0.05/8 = 0.00625 level (two-tailed critical value is 2.73).

ters in all comparisons are independent, their variance estimates were added to obtain test variances. The Bonferroni procedure was used to control for multiple testing by dividing the "experimentwise" error rate of 0.05 by the number of tests to determine the significance level [*Miller*, 1981, pp. 67–70]. Each set of formal comparisons for a specific particle size or sampler was considered to be "an experiment."

In the first set of tests, we investigated, for each sampler, whether calibrations are the same across particle sizes (Table 4). The α and β values were compared to corresponding parameters among each of the uniform particle sizes (three tests each for α and β); and the average of the uniform particle size parameters was compared to the corresponding mixture parameter (one test each for α and β). The eight tests in this experiment dictated a significance level for each test of 0.05/8 = 0.00625, with a corresponding two-tailed critical value of 2.73.

In the second set of tests, we investigated, for each particle size, whether samplers with larger nozzle ratios collect more sediment (Table 5). Samplers with larger ratios of exit to entrance area have higher hydraulic efficiency (Table 1) and are expected to have augmented sediment-trapping efficiency. Again, the test variances were estimated by adding the component independent variance estimates. Six comparisons were made for this experiment with an overall error rate of 0.05, an individual error rate of 0.05/6 = 0.00833, and a one-tailed critical value of 2.39.

The third set of tests was conducted to investigate whether samplers 3 or 6 have higher sediment-trapping efficiency than sampler 1 (Table 6). Sampler 3 has a higher nozzle ratio (3.22 versus 1.40) and sampler 6 has a larger nozzle (152 × 152 mm versus 76 × 76 mm). One experiment was done for each particle size giving an error rate for each test of 0.05/4 = 0.0125 and a one-tailed critical value of 2.24.

In the last set of tests, we compared the calibrations for samplers 1 and 3 to the line $y = x$ (Table 7). In this set of tests the estimated intercepts were compared to 0 and the slopes were compared to 1. Again there was one experiment for each particle size giving an error rate of 0.05/4 = 0.0125 and a two-tailed critical value of 2.50.

TABLE 5. Statistics for Comparing Calibration Parameters for Helley-Smith Samplers With Different Nozzle Ratios for Each of the Four Particle Sizes

Nozzle Ratio Contrast	Parameter	Particle Sizes, mm 2.1	6.5	23.5	Mixture
1.10/1.40	α	0.527	−2.208	−3.291*	−1.443
	β	−1.332	−2.770*	1.490	0.031
1.10/3.22	α	1.431	−5.738*	−3.011*	−1.799
	β	−4.021	−5.677*	−0.872	−2.484*
1.40/3.22	α	0.890	−3.574*	−0.151	0.097
	β	−2.702*	−3.001*	−2.212	−2.451*

*Significant at the 0.05/6 = 0.00833 level (one-tailed critical value is 2.39).

RESULTS AND DISCUSSION

The plots in Figure 5 are inverses of the fitted lines in transformed (i.e., cube root) space with a common scale (in real units) to facilitate comparison. Also shown are the discrimination bands; lines of equal response, $y = x$; and daily means of the transformed Helley-Smith and trap transport data. Each column contains plots for one Helley-Smith sampler and each row for one particle size. Because the calibrations are inverses of the fitted lines, experimental errors are horizontal instead of vertical. The lines extend to the extremes of the individual transformed Helley-Smith transport rates for that calibration.

The back-transformed calibrations have different characteristic features (Figure 9). Back-transforming (i.e., cubing) the linear calibrations generally produces curves (although the curvature is small in these cases because $\hat{\beta}^3 \approx 1$). The calibration lines and most discrimination bands have intercepts much less than one, so cubing tends to yield near-zero intercepts for zero Helley-Smith transport rates. At higher Helley-Smith rates, trap and discrimination values exceed one, making the bands widen rapidly.

Two subsets of plots (Figure 5) have distinctive appear-

TABLE 6. Statistics for Testing Equality of Calibration Parameters Among Helley-Smith Samplers 1, 3, and 6

Sampler Contrast	Parameter	Particle Sizes, mm 2.1	6.5	23.5	Mixture
1/3	α	0.656	2.337*	−0.847	−1.203
	β	0.260	0.446	1.491	2.180
1/6	α	1.131	0.081	−2.490*	−4.578*
	β	−1.172	−8.363*	0.713	2.870*

*Significant at the 0.05/4 = 0.0125 level (one-tailed critical value is 2.24).

TABLE 7. Statistics for Comparing Helley-Smith Samplers 1 and 3 to the Line of Equal Response ($y = x$)

Sampler	Parameter	Particle Sizes, mm 2.1	6.5	23.5	Mixture
1	α	1.125	4.440*	−0.297	0.865
	β	−0.357	1.748	0.743	1.856
3	α	0.291	−0.244	0.910	1.563
	β	−0.795	0.729	−1.384	−1.712

*Significant at the 0.05/4 = 0.0125 level (two-tailed critical value is 2.50).

ances; one consists of samplers 4 and 5 for the 2.1-mm sediment. These calibrations have large slopes (i.e., small fitted values of $\hat{\beta}$). Their discrimination bands were either off scale or too wide to be usable, so they were not plotted. These two samplers had the largest percentages of zero Helley-Smith transport rates for any sampler and particle size, and sampler 4 had a larger proportion of zeros than any trap run. The values of $\hat{\beta}$ are the smallest and the associated estimates of $\hat{\text{Var}}\,[\hat{\beta}]$ the largest for any calibration.

These problems may have resulted partly from the narrow ranges of transformed daily trap means available for these calibrations. Ranges in trap means for the 2.1-mm data are near the narrowest for any particle size and those for samplers 4 and 5 are the smallest for all combinations of sampler and particle size. Their values of $\hat{\sigma}^2$ are large in spite of low mean values; evidently due to the large number of zeros. The estimates of regression slope therefore must be based on highly variable Helley-Smith measurements made over narrow ranges of trap means, conditions under which slopes cannot be estimated with good confidence.

This is reflected by high estimates of $\hat{\text{Var}}\,[\hat{\beta}]$ obtained for combinations of sampler and particle size having narrow ranges of trap means (Figure 10). There is a threshold at a range of about 0.2, below which the variance of the estimated slope rises precipitously. The 2.1-mm and mixture particle sizes had ranges of trap means less than half those for the 6.5- and 23.5-mm data and constituted all 10 of the estimates of Var $[\hat{\beta}]$ above 0.005, with the values for samplers 4 and 5 of the 2.1-mm data both above 0.044.

The other subset of unique plots originally gave negative maximum likelihood estimates of $\hat{\sigma}^2$; samplers 5 and 6 for the 6.5-mm particle size, sampler 4 for the 23.5-mm data, and

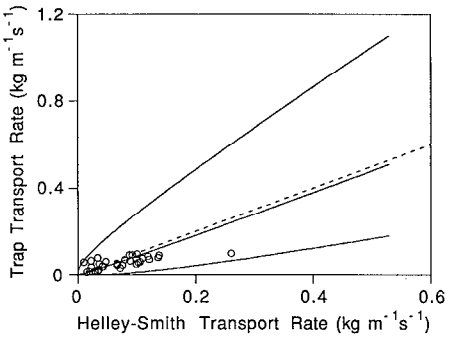

Fig. 9. Typical back-transformed calibration, discrimination bands, line of equal response ($y = x$), and daily means of Helley-Smith and trap data (showing Helley-Smith sampler three and the 6.5-mm particle size).

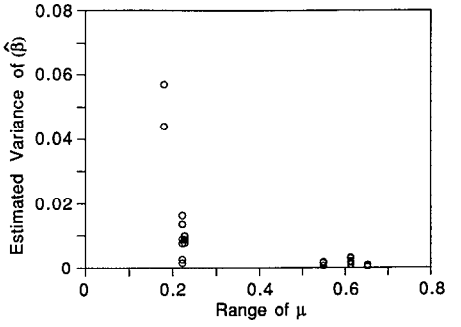

Fig. 10. Estimated variance of $\hat{\beta}$ as a function of the range of daily mean trap transport rates.

samplers 1 and 6 for the mixture data. For these five cases α and β in Table 3 and Figure 5 were estimated by maximizing a likelihood function which assumed that $\sigma^2 = 0$. These calibrations have unusually narrow discrimination bands which result from generally low variances for $\hat{\alpha}$ and $\hat{\beta}$ and $\sigma^2 = 0$, producing small variance estimates for $\hat{x}_0$ in (9). Using $\sigma^2 = 0$ to derive a likelihood function for estimating $\hat{\alpha}$ and $\hat{\beta}$ does not imply zero variance, but is an obvious way to enable estimation using the closest "tenable" value for σ^2.

The reasons for the negative estimates of σ^2 are not clear. The Newton-Raphson method was used to estimate parameters in the maximum likelihood equations because closed-form expressions could not be derived. Therefore characteristic features of the data that might have produced negative variance estimates could not be identified.

Least squares estimates of variance are distributed χ^2, but negative variance estimates for these five cases imply that this variance estimator is not χ^2 (the χ^2 distribution is defined only for nonnegative arguments). Since maximum likelihood estimators are asymptotically normal, some negative estimates might be expected. It is not clear how the experimental or estimation procedures can be changed to avoid negative estimates of σ^2 nor how the estimates for these five cases compare to the other calibrations.

Tests of performance of the Helley-Smith samplers across particle sizes showed no differences for the three samplers (1, 2, and 3) having 76-mm square nozzles (Table 4). That is, the flume data do not justify making a distinction in the responses of these three samplers between the pairs of uniform particle sizes tested or between the average of the uniform sizes and the mixture.

Samplers 4, 5, and 6 show a different pattern, however, with seven significant differences out of 12 comparisons (regarding calibrations as different if either α or β or both are significant (Figure 5)). A likely reason for at least three of these significant values is the low fitted values of $\hat{\beta}$ for samplers 4 and 5 for the 2.1-mm data. Also, four of the five "zero" variances are contained in the calibrations of the three large-nozzle samplers which will tend to produce more sensitive tests. Six of the seven differences are due to slopes and only one to intercepts.

Another way to compare the Helley-Smith samplers is to plot all calibrations on the same graph for each particle size (Figure 8). Particle sizes 6.5 and 23.5 mm show less variation among samplers than do the 2.1-mm and mixture data. It is not clear from these data whether or not this distinction results from the narrow ranges of the 2.1-mm and mixture

data or from actual differences in sampler performance. This question can be answered only by collecting additional data for these particle sizes.

Comparing the response of samplers with different nozzle ratios for each of the particle sizes produced nine out of 12 possible distinctions (Table 5). The 1.10 (samplers 2 and 4) and 3.22 (samplers 1 and 6) nozzle ratios are different for all particle sizes. The 1.10 and 1.40 (samplers 3 and 5) nozzle ratios are significantly different only for the 6.5- and 23.5-mm data. The 1.40 and 3.22 nozzle ratios are different for all except the 23.5-mm data. These tests indicate that Helley-Smith samplers with different nozzle ratios considered as a group generally perform differently. This comparison may be confounded by the possibility that sampler performance differs more due to nozzle size than to nozzle ratio.

Two more specific comparisons were made between samplers 1 and 3, and between 1 and 6, for each particle size (Table 6). The two small-nozzle samplers were compared because sampler 1 is the most widely used sampler and sampler 3 has been recommended based on an earlier analysis of these data (J. V. Skinner, personal communication, 1991). Figure 8 shows that sampler 1, which has the higher nozzle ratio, trapped more sediment for all particle sizes (except at very low transport rates for the mixture); however, only one statistically significant difference was detected.

Sampler 1 was compared to sampler 6 to see if the two standard samplers with identical nozzle ratios but differing nozzle sizes respond similarly. Figure 8 indicates that sampler 6 trapped sediment more efficiently under most conditions. Statistically different responses were shown for particle sizes 6.5 mm, 23.5 mm, and the mixture, but not for the 2.1-mm data.

The last formal tests explored the question of whether or not Helley-Smith samplers 1 and 3 respond the same as the trap if it were placed in the stream. That is, this tested if the calibration line is equal to the line $y = x$. No significant differences between the line $y = x$ were found for sampler 3, but one difference was found for the intercept of the 6.5-mm data for sampler 1 (Table 7).

Several problems in the design of the calibration experiment may have impaired the quality of the results. For example, parameter variances for the model we recommend could be reduced by operating the flume at more levels of stream power over a wider range. More runs of shorter duration could achieve this with no increase in cost. Also, care should be taken to collect sufficient numbers of Helley-Smith measurements on each day of operation (sample sizes of three or four were not uncommon in these data). Finally, extraneous variation would be reduced by operating all samplers from a single station in the center of the flume.

SUMMARY AND CONCLUSIONS

The PMM cannot in general be used to determine the form of a calibration model or estimate its parameters. The process maximizes the correlation possible for a given data set (even turning negative correlation positive) and yields misleading "models" that do not respond to information contained in the data about the true model. The three conditions that have been proposed to justify applying the PMM to the flume bed load data implicitly require that there be no measurement or sampling errors in the trap or sampler data. Virtually all experience, however, indicates that such assumptions are untenable for any real data. Investigation of these conditions for the flume study shows that they cannot hold, even approximately, for this situation in particular. Calibrating bed load samplers is a difficult problem because it is impossible to obtain simultaneous measurements of the samplers and "truth" using presently available technology. However, this does not justify assuming that errors in measurement and sampling do not exist.

A nonstandard regression model (with parameters estimated by maximum likelihood) was developed to account for the physical inability of matching individual trap and Helley-Smith bed load calibration measurements in time or space and the consequent need to allow (curvilinear) calibrations derived from averaged data.

Because the large set of flume measurements was not collected specifically to be used with this model, its usefulness was limited. The ranges of regressor variables were too narrow to define good relationships for most of the 2.1-mm and mixture data. In these cases, variances of the estimated slopes are large and the prediction bands so wide the calibrations are useless. It was difficult to detect significant differences between many of the calibration lines because of large variances of estimated coefficients. Also, five calibrations yielded negative estimates of the measurement error of the Helley-Smith samplers which was set to zero to obtain more reasonable estimates of the other parameters. The tentative conclusions should be viewed in this context.

The three samplers with 76 × 76 mm nozzles performed consistently across particle sizes. Analogous tests on the three large-nozzle (152 × 152 and 152 × 305 mm) samplers identified seven significant differences in the 12 comparisons made. Several of these differences may be attributable to the two poorly fitted calibrations noted above.

Comparisons between samplers with differing nozzle ratios yielded nine significant differences out of 12 tests. These comparisons may also have been influenced by the two poorly fitted cases as well as the fact that each nozzle ratio was represented by a large and a small sampler.

In two comparisons involving the standard samplers, nozzle size seemed more important than nozzle ratio in determining sediment-trapping efficiency. The sampler with the larger nozzle collected sediment more efficiently.

Future attempts to collect calibration data for bed load samplers can benefit from the lessons of this project. Although not presently anticipated, the ideal solution would be to develop a procedure to measure true sediment flux "simultaneously" with sampler measurements to obtain truly matched individual observations. Then standard transformation, calibration, and inverse estimation techniques could be used to specify calibration models and estimate parameters and variation.

Barring this solution, we recommend a more complex model that matches sampler measurements with time-averaged determinations of true transport rate. The model incorporates a term which explains part of the variation in sampler measurements using the variation in the data that compose each true mean. In either case, there are several recommendations that will help ensure a satisfactory outcome of the calibration.

First, decide (in detail) on the analysis before collecting data and structure the experimental procedures accordingly. The importance of this obvious, but frequently ignored step,

cannot be overemphasized; creating a model to fit data collected without anticipating the details of analysis virtually always requires substantial compromises.

Second, use a model based on established statistical principles to better enable analysts to understand its behavior, to specify model form, and to obtain valid estimates of parameters.

Third, ascertain the source of idiosyncracies in the data, such as outliers or large numbers of zero transport rates, to obviate having to remove these data arbitrarily from the data sets.

Fourth, collect data for enough hydraulic conditions to produce a range of bed load transport rates adequate for precise estimation of model parameters and spanning the range expected under field conditions (e.g., emphasize more levels of stream power with fewer days at each level).

Fifth, ensure that the range of transport rates is covered approximately uniformly to enable proper assessment of model fit.

Sixth, collect enough data in each hydraulic condition to counter the high variability inherent in measuring and sampling bed load transport and to precisely estimate model parameters. This can be facilitated by limiting the scope of the study to ensure that the specified goals can be achieved.

Acknowledgments. The authors thank David Hubbell and Herbert Stevens for their willing and invaluable assistance in helping us understand the philosophy and details of data collection for the bed load sampler study. We also thank Sylvia Mori and James Baldwin for extensive statistical discussions, assistance, and suggestions associated with analyzing these data and Elizabeth Maxwell who first studied the problems with the probability matching method in 1985.

REFERENCES

Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, pp. 492–497, MIT Press, Cambridge, Mass., 1975.

Box, G. E. P., and D. R. Cox, An analysis of transformations, *J. R. Stat. Soc., Ser. B, 26*, 211–252, 1964.

Chambers, J. M., W. S. Cleveland, B. Kleiner, and P. A. Tukey, *Graphical Methods of Data Analysis*, pp. 53 and 56, Duxbury Press, Boston, Mass., 1983.

Chapra, S. C., and R. P. Canale, *Numerical Methods for Engineers*, 2nd ed., pp. 243–245, McGraw-Hill, New York, 1988.

de Vries, M., On measuring discharge and sediment transport, *Publ. 106*, 15 pp., Delft Hydraul. Lab., The Netherlands, 1973.

Emmett, W. W., A field calibration of the sediment-trapping characteristics of the Helley-Smith bedload sampler, *U.S. Geol. Surv. Prof. Pap., 1139*, 1980.

Gibbons, J. D., *Nonparametric Statistical Inference*, pp. 73–74, McGraw-Hill, New York, 1971.

Gomez, B., R. L. Naff, and D. W. Hubbell, Temporal variations in bedload transport rates associated with the migration of bedforms, *Earth Surf. Processes Landforms, 14*(2), 135–136, 1989.

Helley, E. J., and W. Smith, Development and calibration of a pressure-difference bedload sampler, *U.S. Geol. Surv. Open File Rep.*, 18 pp., 1971.

Hoaglin, D. C., F. Mosteller, and J. W. Tukey, *Understanding Robust and Exploratory Data Analysis*, pp. 98–112, John Wiley, New York, 1983.

Hogg, R. V., and A. T. Craig, *Introduction to Mathematical Statistics*, 3rd ed., pp. 254–257, Macmillan, New York, 1970.

Hubbell, D. W., and H. H. Stevens, Factors affecting accuracy of bedload sampling, *Proc. Fed. Inter Agency Sediment Conf. 4th*, *2*(4), 20–29, 1986.

Hubbell, D. W., H. H. Stevens, J. V. Skinner, and J. P. Beverage, Recent refinements in calibrating bedload samplers, in *Water Forum '81*, vol. 1, pp. 128–140, American Society of Civil Engineers, New York, 1981.

Hubbell, D. W., H. H. Stevens, J. V. Skinner, and J. P. Beverage, New approach to calibrating bed load samplers, *J. Hydraul. Eng., 111*, 677–694, 1985.

Hubbell, D. W., H. H. Stevens, J. V. Skinner, and J. P. Beverage, Laboratory data on coarse-sediment transport for bedload sampler calibrations, *U.S. Geol. Surv. Water Supply Pap., 2299*, 1987.

Kendall, M. G., and A. Stuart, *The Advanced Theory of Statistics*, 2nd ed., vol. 2, p. 55, Charles Griffin, London, 1967.

Miller, R. G., *Simultaneous Statistical Inference*, Springer-Verlag, New York, 1981.

Weisberg, S., *Applied Linear Regression*, pp. 9, 10, and 19, John Wiley, New York, 1985.

J. Lewis and R. B. Thomas, Pacific Southwest Research Station, U.S. Department of Agriculture, 1700 Bayview Street, Arcata, CA 95521.

(Received May 21, 1992;
revised September 11, 1992;
accepted September 24, 1992.)