A Comparison of Load Estimates Using Total Suspended Solids and Suspended-Sediment Concentration Data

G. Douglas Glysson,* John R. Gray,** and Gregory E. Schwarz ***

* U.S. Geological Survey, 412 National Center, Reston, VA 20192; PH (703) 648-5019;
FAX (703) 648-5722; e-mail gglysson@usgs.gov
** U.S. Geological Survey, 415 National Center, Reston, VA 20192; PH (703) 648-5318;
FAX (703) 648-5722; e-mail jrgray@usgs.gov;
*** U.S. Geological Survey, 413 National Center, Reston, VA 20192; PH (703) 648-5718;
FAX (703) 648-5295; e-mail gschwarz@usgs.gov;

Abstract

This paper presents the results to-date from a continuing investigation into the differences between total suspended solids (TSS) and suspended-sediment concentration (SSC) data and the ramifications of using each type of data to estimate sediment loads. It compares estimates of annual suspended-sediment loads that were made using regression equations developed from paired TSS and SSC data, to annual loads computed by the U.S. Geological Survey (USGS) using traditional techniques and SSC data. Load estimates were compared for 10 stations where sufficient TSS and SSC paired data were available to develop sediment-transport curves for the same time period that daily suspended-sediment records were available. Results of these analyses indicate that as the time frame over which the estimates were made increases, the overall errors associated with the estimates decreases with respect to loads computed using traditional USGS techniques. Using SSC data to compute loads tends to produce estimates closer to those computed by traditional techniques than those computed from TSS data. Loads computed from TSS data tend to be negatively biased with respect to those computed by traditional USGS techniques.

Introduction

The U.S. Environmental Protection Agency (2000) identifies fluvial sediment as the most widespread pollutant in the Nation's rivers and streams, affecting aquatic habitat, drinking water treatment processes, and recreational uses of rivers, lakes, and estuaries. Two laboratory analytical methods – suspended-sediment concentration (SSC), and total suspended solids (TSS) – are predominantly used to quantify concentrations of suspended material in surface waters of the United States. The analytical methods differ. The SSC method analyzes all of the sediment and the mass of the water-sediment mixture within the sample to derive concentration values (ASTM, 2000). The TSS method, which originated as a wastewater analytical method, analyzes a subsample of the original sample to compute concentrations of suspended solids (American Public Health Association, American Water Works Association, and Water Pollution Control Federation, 1995). In practice, TSS data are produced by a variety of methods and equipment. An analysis of quality-control data (Gordon and others, 2000; U.S. Geological Survey, 1999) indicated greater variances associated with the TSS data than with SSC data produced by the same laboratory, and TSS data were commonly negatively biased with respect to SSC data. Gray and others (2000) found that SSC values tend to exceed TSS values from concomitantly collected samples, particularly when the percentage of sand-size material exceeds about a quarter of the

sample sediment mass. Gray and others (2000), Glysson and others (2000), and the U.S. Geological Survey (2000a) identify the TSS analytical method as fundamentally unreliable for use in quantifying concentrations of sediment in surface waters.

Bias in the relation between TSS and SSC data has important ramifications for estimating sediment discharges. Sediment discharge increases when the product of water discharge and SSC increases. Additionally, the mobility of coarse material tends to increase with the larger flow velocities associated with larger discharges. Because of the strong tendency for SSC to exceed TSS at larger values of SSC (Gray and others, 2000), suspended solid-phase discharges calculated from TSS data will tend to be underestimated with respect to those calculated from the product of the same water-discharge time series and SSC data.

This paper demonstrates that load computation estimates based on TSS data and sediment transport curves can differ substantially from those based on concurrently collected SSC data and the same water-discharge time series. It also compares the load estimates based on TSS, and on SSC data, to those measured and computed by graphical computational techniques traditionally used by the USGS (Porterfield, 1972).

Data

Concentration samples. A total of 14,466 data pairs analyzed using the SSC (USGS parameter code 80154) and TSS (USGS parameter code 00530) methods were retrieved from the electronic files of the USGS (U.S. Geological Survey, 2000b). Data were available from 48 States and Puerto Rico.

Samples collected sequentially in-stream may have different concentrations and size characteristics of solid-phase material. This may be due to natural variations in amounts and the composition of solid-phase material in transport, and to variance and (or) bias introduced by sampling procedures. Likewise, a subsample may contain an amount and size distribution of sediment atypical of the original sample. However, any bias in individual sediment-concentration and size-distribution data resulting from in-stream variations would likely occur randomly among the SSC-TSS data pairs (Glysson and others, 2000). Although in-stream sampling or subsequent field processing procedures may have increased the variance of the derivative data, there is no evidence that these procedures are responsible for bias in the SSC-TSS relation.

Daily suspended-sediment loads. The USGS Daily Suspended-Sediment Database (U.S. Geological Survey, 2000c) contains sediment-load records for 1,593 stations through September 1994 that have an average period of record of 5.3 years. The period of record for individual stations ranges from one day to 44 years. Thirteen percent of the stations have 10 years or more of record. These daily records were computed using the standard USGS techniques described by Porterfield (1972); each year of record normally has 200 to 300 analytical results from samples available for the computation. The daily records from this database are referred to hereafter as "actual" suspended-sediment loads produced by "traditional USGS techniques." Although there are no estimates of bias and variance associated with sediment-load data computed by techniques described by Porterfield (1972), these records are generally considered to be the most accurate sediment-load data available for the stations at which they were

collected. It is assumed in the remainder of this paper that, for comparison purposes, sediment loads computed from traditional USGS techniques for a station are the true loads.

Criteria for selection of stations used in this analysis. Inclusion of data for a station in this analysis required that all of the following conditions be met: (1) at least 10 or more paired SSC-TSS data pairs available for the station; (2) at least 3 years of continuous daily suspended-sediment load data were available; and (3) the samples were collected during the period of daily load record. Only the 10 stations listed in Table 1 met these criteria and were used in this analysis. Some of these stations had additional periods of daily load records not used in this analysis because the unused period of record did not contain any SSC-TSS data pairs and was not continuous with the period of record listed in table 1. Also, some stations had additional paired SSC-TSS that were not used in this analysis because they were collected outside of the period of daily load record.

Table 1. List of stations used, number of sample pairs used at each station, and the period of record for samples and daily suspended-sediment loads

Station ID	Station Name	Number of Sample Pairs	Sampling Period (month/year)		Period of Daily Record Used (month/day/year)	
			Begin	End	Begin	End
01463500	Delaware R. @ Trenton, NJ	40	3/1974	10/1978	10/1/1950	9/30/1981
05325000	Minnesota R. @ Mankato, MN	52	3/1994	9/1995	10/1/1967	9/30/1995
05406470	Brewery Cr. @ Cross Plains, WI	93	10/1991	6/1994	10/1/1991	9/30/1994
05594100	Kaskaskia R. nr Venedy Station, IL	10	10/1980	10/1985	10/1/1980	9/30/1988
05599500	Big Muddy R. nr Murphysboro, IL	13	10/1980	9/1988	10/1/1980	9/30/1988
06214500	Yellowstone R. @ Billings, MT	12	10/1976	9/1981	10/1/1976	9/30/1981
06308500	Tongue R. @ Milles City, MT	14	10/1977	8/1982	10/1/1977	9/30/1985
08313000	Rio Grande @ Otowi Bridge, NM	37	10/1974	8/1980	10/1/1955	9/30/1989
09368000	San Juan R. @ Shiprock, NM	36	10/1973	9/1980	10/1/1950	9/30/1986
12510500	Yakima R. @ Krona, WA	35	10/1977	9/1980	10/1/1977	9/30/1980

Analysis of the Data

Regression Analysis: A line representing the relation between sediment and water discharge is commonly referred to as a sediment-transport curve. Transport curves delineated by a single straight line fitted through a set of points will not necessarily define an accurate relation between the variables (Glysson, 1987). The following equation was used to define the relation between concentration data and daily water-discharge data:

$$\text{Log SL} = b_0\text{int} + b_1\text{ldflow} + b_2\cos 2\pi t + b_3\sin 2\pi t + b_4\text{ldfhi} + e,$$

where:
SL = suspended-sediment load in tons/day; b_x = regression coefficients for the indicated factor in the equation; $b_0$int = y intercept; ldflow = natural logarithm of the water discharge; t = decimal time of the year in the form of yr. (day/365); ldfhi = the amount the natural logarithm of flow differs from the natural logarithm of a breakpoint discharge (see next paragraph) where the transport curve may have a break in slope. Negative differences are set to zero; and e = a random error term.

For this analysis, SSC values were plotted with their corresponding water-discharge values on logarithmic scales and visually examined to determine if two ordinary least-squares (OLS) regression lines of differing slope would be an improvement over a single-line model. If two OLS regression lines were deemed appropriate, the water discharge at which the lines intersected, otherwise referred to as the "break in slope," was estimated. The discharge at this break in slope was used in the bootstrapping program (defined below in the "Load Estimation Method" section) and is represented by the ldfhi variable. The magnitude of the break in slope of the regression lines was determined by regression analysis. Table 2 is a summary of the results of the regression analysis.

Sediment-transport curves can exhibit seasonal trends (Glysson, 1987). To investigate the effect that seasonality might have on the relation of sediment concentration to water discharge for the 10 stations used in this study, sine and cosine functions were added to the regression equation. Data for all stations were estimated using ordinary least squares regression analyses for models with and without seasonal cycles. A standard F test showed that for several stations, the single cycle sine and cosine functions were significant.

Table 2. Summary of regression coefficients that produced the lowest percent error in estimating suspended-sediment loads for the period of record and the root mean square of the error (rmse) for the estimates

Station ID	Name	Type	b_0	b_1	b_2	b_3	b_4	rmse[1]
01463500	Delaware R. @ Trenton, NJ	SSC	-11.132	1.443	-0.851	-0.344	0	0.95
		TSS	-4.059	0.678	0	0	0	1.02
05325000	Minnesota R. @ Mankato, MN	SSC	0.650	0.543	0	0	-0.521	0.44
		TSS	-7.822	10449	0.333	-0.596	-1.703	0.54
05406470	Brewery Cr. @ Cross Plains, WI	SSC	3.800	0.590	0	0	0.460	0.98
		TSS	3.117	0.953	0	0	-0.115	1.21
05594100	Kaskaskia R. nr Venedy Station, IL	SSC	2.664	0.263	0	0	0	0.52
		TSS	2.881	0.232	0	0	0	0.56
05599500	Big Muddy R. nr Murphysboro, IL	SSC	4.210	-0.011	0	0	0	0.54
		TSS	4.512	-0.097	-0.779	-0.148	0	0.32
06214500	Yellowstone R. @ Billings, MT	SSC	-0.932	0.468	0	0	3.476	0.94
		TSS	-4.391	0.817	0	0	1.840	0.96
06308500	Tongue R. @ Milles City, MT	SSC	2.190	0.358	0	0	0	1.09
		TSS	1.441	0.289	-1.910	-0.284	0	0.83
08313000	Rio Grande @ Otowi Bridge, NM	SSC	2.909	0.394	0	0	0	1.25
		TSS	2.182	0.337	0	0	0	1.13
09368000	San Juan R. @ Shiprock, NM	SSC	4.330	0.342	0.917	0.728	0	1.17
		TSS	5.114	0.084	0.574	0.709	0	1.27
12510500	Yakima R. @ Krona, WA	SSC	6.282	-0.443	0	0	1.900	0.66
		TSS	8.793	-0.840	0	0	2.421	0.89

[1] rmse = root mean square of the error (error, in percent, approximately equal to one half the 95% confidence level expressed as a plus/minus percentage of the load.)

Load Estimation Method

Daily and annual sediment loads were estimated using bootstrap analysis (Efron, 1982), a robust estimation program that does not assume any particular distribution for the error term of the regression model. The only limitation in its use is that the error term must be serially independent, identically distributed, and independent of the explanatory variables. The program is implemented by forming repetitive resamples of the N available observations. Each resample is formed by randomly selecting, with replacement, N observations from the original sample. Thus, for the first resample, the first observation may appear twice and the second observation may be excluded; for the second resample, the first observation may appear once and the second observation may appear twice. A total of B resamples are selected in this manner, representing B iterations of the bootstrap program. For each bootstrap iteration, the regression model is estimated and the coefficient estimates and estimated errors are used to simulate load for each day of the flow record. One-hundred iterations were used in this analysis for all stations. Simulation of a daily load value is based on the log-linear regression equation; using the bootstrap iteration's coefficient estimates and randomly selecting one of the N estimated errors. The resulting simulated logarithmic load value is then exponentiated and stored. Simulated daily values for a water year are aggregated to form a bootstrap iteration estimate of the annual load. Repeating this process for all B bootstrap iterations results in B estimates of each daily and annual load value in the flow record. The average and standard deviation of these B estimates is the bootstrap mean and standard deviation estimate of daily and annual load.

Discussion of Results

Inaccuracies of regression analyses: Estimates of suspended-sediment loads based on regression analyses are subject to significant errors (Glysson, 1987). Because of the nature of sediment transport in open channels, there can be a large range in sediment concentrations at any given discharge. The fewer the number of concentration values available to define this range, the larger the potential errors can be.

Progressively better results in the estimates as the time period increases: The magnitude of variations resulting from the use of regression analysis to estimate sediment loads decreased substantially with respect to those computed by traditional USGS techniques as the time frame of the estimated value increases. For example, errors between daily-sediment loads computed by regression versus traditional USGS techniques at the Rio Grande at Otowi Bridge station were as large as 4000%, while the maximum error in the estimation of an annual load was 526%. However, the error in the estimate of the total suspended-sediment load for the 34 years of record at this station is only 38%. Table 3 shows the errors in the estimate of the suspended-sediment loads for the 10 stations used in this analysis.

Comparison of total suspended-sediment load estimates using SSC and TSS data: The results in Table 3 show that at 6 of the 10 stations analyzed, the error in total estimated load using TSS data for the period of record was at least 30% smaller than their respective actual loads. The other four stations (Brewery Creek at Cross Plains, Wisconsin; Kaskaskia River near Venedy Station, Illinois; Tongue River at Milles City, Montana; and Yakima River at Krona, Washington) had comparable results (differences in errors less than 15%) when either the TSS or SSC data were used. On the basis of this

comparison, if one was to use the TSS data to estimate the long-term suspended-sediment load for a station, one might expect the estimates to be substantially lower (biased 20% or more low) than actual loads for about 70% of the stations. However, if SSC data were used, loads for only about 30% of the sites might be expected biased 20% or more lower than actual loads.

Ten stations represent an inadequate sample size to make an accurate prediction of the percentage of time each type of data will give an acceptable result. This analysis does give an indication, however, that the chances of obtaining a better estimate is significantly greater when using SSC then when using TSS data.

At 6 of the 10 stations included in the analysis, errors in the total estimated load for the period of record were equal to or greater than 40% when the TSS data were used, compared to an error of such magnitude at only one station when the SSC data were used. As a hypothetical example, if one used the results of this analysis based on TSS data to estimate the reduction in loads for a Total Maximum Daily Load (TMDL) study, an estimate that was initially biased 40% low might be unacceptable. If the post-remediation sediment load that was needed to meet the TMDL criteria was, for example, a 30% reduction in the sediment load from the estimate made, the best management practices imposed on the watershed would be targeted to that percentage. When monitoring in the TMDL reach begins and the SSC analysis is used – as it should be to accurately describe the suspended-sediment concentration in an open channel (Gray and other 2000; U.S. Geological Survey, 2000a) – the results of the monitoring program would indicate an increase in the sediment load (assuming a similar hydrologic regime and land use existing at the time that the historical data were collected). This might cause the investigator to erroneously conclude that the best management practices implemented to remediate the TMDL reach were having a detrimental rather than beneficial effect on the sediment-transport problem.

Can one tell if the TSS data will give as good or better estimates of suspended-sediment loads without having traditional daily suspended-sediment data with which to compare? The answer to this question is that apparently one cannot tell if the TSS data will give an accurate estimate simply by comparing the relation between the TSS and SSC data. Figure 1 shows plots of the paired SSC and TSS data for four of the stations used in this analysis. The plots of the Delaware River, and Brewery Creek stations show a fairly good relation and little difference between the TSS and SSC concentration pairs. The errors in the total estimated load for the Brewery Creek station are relatively small (TSS = 9% and SSC = -3%), but the estimate of the total sediment load for the Delaware River station is –40% when using the TSS data and only -5% when the SSC data were used. Additionally, total loads using TSS data for the San Juan and Big Muddy River stations were significantly underestimated (-80% and –65%, respectively). A review of the plots of the data in Figure 1 shows the regression line for the Big Muddy River station, which was fitted to the paired data, converges with the line of equal value as the concentrations increase but diverges for the San Juan River station plot. One might construe that the converging relation might give a better estimate because the differences between the TSS and SSC values lessen as the concentration increases. This was not true, at least at the Big Muddy River station.

Summary

1. Use of regression analysis to develop sediment-transport curves for estimating suspended-sediment loads can result in substantial errors. The absolute value of errors in this study ranges from as large as 4000% for the estimation of a daily load to 2% for the estimation of the sum of the loads for the period of record. In all cases, differences found between the actual suspended-sediment loads computed by traditional USGS techniques and associated estimated loads decreased as the time period over which the loads were estimated increased.

    A well-defined, carefully constructed, and judiciously applied sediment-transport curve can be a useful tool for estimating sediment loads. However, because of the potential large errors associated with using regression analysis for sediment-load estimates, loads calculated from transport curves should not be considered a substitute for daily-sediment records.

2. The magnitude of the errors in load estimates from the two data types with respect to traditional computational techniques is at least partly a consequence of the relatively few observations used to define the regression relations. However, the relative errors between loads computed by transport curves defined by SSC versus TSS are indicative of the relative errors associated with the two analytical methods.

3. Load estimates using SSC data tend to have smaller errors than those for which TSS data were used. Six of the 10 sites included in the analysis indicated errors in the sum of the loads larger than 40% when the TSS data were used, compared to only one when the SSC data were used. No stations had the errors in the sum of loads using TSS data significantly smaller than those using SSC data.

4. No simple, straightforward means for comparing SSC-TSS paired data has been identified to determine if the TSS data will give as good or better estimate of the suspended-sediment load.

References

American Public Health Association, American Water Works Association, and Water Pollution Control Federation (1995), *Standard Methods for the Examination of Water and Wastewater* (19th ed.): Washington, D.C., American Public Health Association, variously paged.
American Society for Testing and Materials (ASTM), 2000, *Standard Test Method for Determining Sediment Concentration in Water Samples: D 3977-97*, vol. 11.02, Water (II), pp. 395-400.
Efron, B. (1982) *The Jackknife, the Bootstrap, and other Resampling Plans*, in CBMS – National Science Foundation Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, PA.
Gray, John R., Glysson, G. Douglas, Turcios, Lisa M., and Schwarz, Gregory E., (2000) *Comparability of Suspended-Sediment Concentration and Total Suspended Solids Data*, U.S. Geological Survey Water-Resources Investigations Report 00-4191, 14 p.
Glysson, G. Douglas (1987) *Sediment Transport Curves,* U.S. Geological Survey Open-File Report 87-218, 47 p.

Glysson, G.D., Gray, J.R., and Conge, L.M., 2000, *Adjustment of Total Suspended Solids Data for Use in Sediment Studies*: Proceedings, ASCE's 2000 Joint Conference on Water Resources Engineering and Water resources Planning and Management, July 31 – August 2, 2000, Minneapolis, Minn., 10 p.

Gordon, J.D., Newland, C.A., and Garliardi, S.T., 2000, *Laboratory Performance in Sediment Laboratory Quality-Assurance Project, 1996-98*: U.S. Geological Survey Water-Resources Investigations Report 99-4184, 69 p.

Porterfield, George (1972) *Computation of Fluvial-Sediment Discharge*, Techniques of Water-Resources Investigations of the United States Geological Survey, Chapter C3, Book 3, 66 p.

U.S. Environmental Protection Agency, 2000, *National Water Quality Inventory, 1998 Report to Congress*, EPA 841-R-00-001, 413 p.

U.S. Geological Survey (1999), *Sediment Laboratory Quality Assurance Project*, accessed January 27, 2000, at URL http://sedserv.cr.usgs.gov.

U.S. Geological Survey (2000a), *USGS Policy on Collection and Use of Total Suspended Solids Data*: Office of Surface Water and Office of Water Quality Memorandum No. 2001.03, accessed January 9, 2001, at http://water.usgs.gov/admin/memo/QW/qw01.03.sw01.03--Collection.and.Use.of.Total.html.

U.S. Geological Survey (2000b), *Water Resources Data for USA*, accessed September 15, 2000, at http://water.usgs.gov/nwis/.

U.S. Geological Survey (2000c), *Suspended Sediment Database: daily values of suspended sediment and ancillary data*, accessed September 15, 2000, at http://webserver.cr.usgs.gov/sediment/.

Table 3 - Summary of years of record and errors* in the estimations of suspended-sediment loads (all errors are in percent)

Site ID	Name	Years of Record	Maximum Annual Error		Minimum Annual Error		Median Annual Error		Mean Annual Error		Error in Total Est. Load for Period of Record**	
			SSC	TSS	SSC	TSS	SSC	TSS	SSC	TSS	SSC	TSS
01463500	Delaware R. @ Trenton, NJ	32	126	-20	-73	-85	-30	-57	-22	-57	-5	-40
05325000	Minnesota R. @ Mankato, MN	28	40	17	-57	-87	-2	-53	-8	-53	-8	-47
05406470	Brewery Cr. @ Cross Plains, WI	4	60	56	-36	-38	-4	7	4	8	-3	9
05594100	Kaskaskia R. nr Venedy Station, IL	8	28	23	-49	-50	11	7	2	-1	6	2
05599500	Big Muddy R. nr Murphysboro, IL	8	34	-48	-60	-77	-18	-63	-14	-64	-13	-65
06214500	Yellowstone R. @ Billings, MT	5	55	-23	-35	-41	-25	-38	-3	-35	8	-34
06308500	Tongue R. @ Milles City, MT	8	-47	-68	-87	-95	-63	-76	-63	-78	-68	-79
08313000	Rio Grande @ Otowi Bridge, NM	34	526	104	-91	-97	-54	-83	-4	-67	-38	-78
09368000	San Juan R. @ Shiprock, NM	31	259	10	-91	-97	-30	-78	-4	-69	-38	-80
12510500	Yakima R. @ Krona, WA	3	13	6	-32	-44	-7	-35	-0	-25	-8	-19

* Error = 100(estimated load – measured load)/ measure load.
** The sum for the period of record of the measured load and the estimated load were used in this calculation.

Figure 1. Plots of SSC versus TSS paired data and ordinary least square regression lines for 4 stations (mg/L = milligrams per Liter; TSS = Total Suspended Solids; SSC = Suspended Sediment Concentration)