WATER RESOURCES RESEARCH, VOL. 28, NO. 9, PAGES 2353–2363, SEPTEMBER 1992

The Validity of a Simple Statistical Model for Estimating Fluvial Constituent Loads: An Empirical Study Involving Nutrient Loads Entering Chesapeake Bay

TIMOTHY A. COHN,[1] DANA L. CAULDER,[2] EDWARD J. GILROY,[1] LINDA D. ZYNJUK,[3] AND ROBERT M. SUMMERS[4]

We consider the appropriateness of "rating curves" and other log linear models to estimate the fluvial transport of nutrients. Split-sample studies using data from tributaries to the Chesapeake Bay reveal that a minimum variance unbiased estimator (MVUE), based on a simple log linear model, provides satisfactory load estimates, even in some cases where the model exhibited significant lack of fit. For total nitrogen (TN) the average difference between the MVUE estimates and the observed loads ranges from −8% to +2% at the four sites. The corresponding range for total phosphorus (TP) is −6% to +5%. None of these differences is statistically significant. The observed variability of the MVUE load estimates for TN and TP, which ranges from 7% to 25% depending on the case, is accurately predicted by statistical theory.

INTRODUCTION

There has been increasing concern during the past several decades about changes in the quality of the water in many of the nation's estuaries, lakes and rivers. In some cases, degradation of water quality has caused measurable economic damage in terms of loss or reduction of both fisheries and recreational opportunities. The changes in water quality are often attributed to excess nutrients, principally compounds containing phosphorus and nitrogen, entering the water body. As a consequence, many remediation programs focus on limiting the influx of nutrients, and this has led to a demand for accurate estimates of the loads of nutrients entering water bodies from tributary streams.

In general, simple random sampling designs are not efficient for measuring most nutrient loads. It costs between $100 and $1000 to collect and analyze a single water sample, and thus it is desirable to design sampling schemes to minimize the number of samples that need to be collected. In particular, researchers often employ log linear models relating the nutrient concentrations to surrogate variables which can be measured less expensively, usually discharge and time. In some cases, these two surrogate variables may explain more than half the total variability in constituent transport rates, allowing one to reduce substantially the number of samples required to attain a given level of precision in load estimates.

Log linear models are attractive for several other reasons: The statistical properties of the standard estimators are well understood and they are easy to apply. In addition, as *Colby* [1956, p. 155] points out,

> . . . daily sediment discharges computed from sediment rating curves may be more accurate than those that are computed from daily samples . . .

This is because the daily observations are subject to the sampling error of a single observation, while the model estimates are based on the entire data set.

This paper considers three simple load estimation methods based on log linear models. They are compared in a repeated split-sample study employing nutrient data from the major tributaries to Chesapeake Bay. Three forms of nitrogen and three forms of phosphorus are considered at each of four sites. It is found that the constituent data do not generally conform exactly to the postulated log linear model: The data always exhibit statistically significant lack of fit, and one occasionally sees substantial serial correlation in the residuals (see *Draper and Smith* [1981] for discussion). However, load estimators based on these models appear to be relatively insensitive to the violations of the hypotheses. In particular, a minimum variance unbiased estimator, based on the assumption of a log linear model, provides good estimates of the loads despite the violations of the assumptions. Furthermore, the sampling properties of the estimates correspond well to those predicted by log linear model theory.

LITERATURE REVIEW

The hydrology literature contains many papers related to methods for estimating sediment loads carried by streams [*Miller*, 1951; *Colby*, 1956; *Brown et al.*, 1970; *Walling*, 1977; *Verhoff et al.*, 1980; *Thomas*, 1985, 1988; *Ferguson*, 1986; *Koch and Smillie*, 1986; *Thompson et al.*, 1987]. These load estimation techniques are, in general, based on the "rating curve," which is usually equivalent to a simple log linear model. It is assumed that the concentration of sediment, C, is related to discharge according to the following equation:

$$C = \exp(\beta_0 + \beta_1 \ln[Q] + \varepsilon) \quad (1)$$

where Q is discharge, β_0 and β_1 are parameters to be estimated, and ε is an independent random error often assumed to be normally distributed with zero mean and variance σ_ε^2. The errors might be related to sample collection, preservation, and analysis, or to natural variability due to processes not represented in the model. The corresponding instantaneous load [see *Glysson*, 1987], L, is given by:

[1]U.S. Geological Survey, Reston, Virginia.
[2]Research Science Institute, McLean, Virginia.
[3]U.S. Geological Survey, Towson, Maryland.
[4]Maryland Department of the Environment, Baltimore.

Paper number 92WR01008
0043-1397/92/92WR-01008$05.00

$$L = KQ \exp(\beta_0 + \beta_1 \ln[Q] + \varepsilon) = \exp(\beta_0' + \{\beta_1 + 1\} \ln[Q] + \varepsilon) \quad (2)$$

where K is a conversion factor and $\beta_0' = \beta_0 + \ln[K]$.

The parameters, β_0 and β_1, are usually estimated "by eye" or by linear regression–least squares. Given the least squares parameter estimates, denoted $\hat{\beta}_0$ and $\hat{\beta}_1$, (2) defines a load estimator for a given flow Q_i:

$$\hat{L}_{RC,i} \equiv \exp(\hat{\beta}_0' + \{\hat{\beta}_1 + 1\} \ln[Q_i]) \quad (3)$$

The subscript RC denotes "rating curve." It will be convenient later to employ matrix notation, in which $\hat{L}_{RC,i}$ can be expressed as $\exp(\mathbf{X} \cdot \hat{\boldsymbol{\beta}})$, where

$$\hat{\mathbf{L}} = \begin{bmatrix} \hat{L}_{RC,1} \\ \hat{L}_{RC,2} \\ \cdots \\ \hat{L}_{RC,N} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & \ln[Q_1] \\ 1 & \ln[Q_2] \\ \cdots & \cdots \\ 1 & \ln[Q_N] \end{bmatrix} \quad \hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0' \\ \hat{\beta}_{1+1} \end{bmatrix} \quad (4)$$

The Validity of Log Linear Models: Some Concerns

Since *Miller* [1951], many researchers have made use of (2) to estimate sediment loads. Many flaws have been identified, including retransformation bias [*Ferguson*, 1986; *Koch and Smillie*, 1986]; lack of fit due to missing variables [*Miller*, 1951; *Colby*, 1956; *Walling*, 1977; *Thomas*, 1988]; and nonnormality of the error distribution [*Thomas*, 1988]. It has been noted frequently that the validity of the models is uncertain and must be tested empirically for each case. *Thomas* [1988, pp. 512–513] reviews some of the problems, and in his example concludes that

> The basic problem is that the model is not adequately specified. There is certainly a positive correlation between water discharge and simultaneously collected suspended sediment concentration in most streams. There is not enough information in the discharge, however, to predict concentration in the commonly used simple linear regression context . . . There is also doubt about exactly what a rating curve represents. Rating curve predictions of suspended sediment discharge are evidently dependent on information not available in the explanatory variable.

Thomas [1985, 1988] derives and recommends a sampling design and estimation method, called SALT, whose statistical properties do not depend on the validity of log linear models. The SALT method is inherently unbiased, and is shown to work well for modeling sediment yields in small watersheds. However, the method requires programmable automated samplers, and it has not, to our knowledge, been widely applied.

Nonetheless, because of the convenience of the log linear model, some researchers continue to develop additional theory assuming the log linear model provides a reasonable description of the natural system [e.g., *Cohn et al.*, 1989; *Gilroy et al.*, 1990], and many researchers employ it. In fact, log linear models seem to be increasing in popularity and are often applied to constituents other than sediment.

Bias Adjustments

Ferguson [1986] and *Koch and Smillie* [1986] point out that the basic rating curve estimator will, in many circumstances, tend to underestimate the true loads. That is,

$$E[\hat{L}_{RC}] \approx E[\exp(\mathbf{X} \cdot \hat{\boldsymbol{\beta}})] \quad (5)$$

while the expected value of the load is equal to

$$E[L] = \exp(\mathbf{X} \cdot \boldsymbol{\beta}) \exp(\sigma_\varepsilon^2/2) \quad (6)$$

Ferguson [1986] and *Koch and Smillie* [1986] point out that this retransformation bias may lead to underestimation of sediment loads by as much as 50%. They propose a new estimator,

$$\hat{L}_{QMLE} \equiv \exp(\mathbf{X} \cdot \hat{\boldsymbol{\beta}} + s^2/2) = \hat{\mathbf{L}}_{RC} \exp(s^2/2) \quad (7)$$

where s^2 is the estimated variance of the residuals following regression. The subscript QMLE denotes "quasi-maximum likelihood estimator." *Cohn et al.* [1989] show that this is upwardly biased, in some cases substantially so; Cohn et al. propose employing a minimum variance unbiased estimator,

$$\hat{L}_{MVUE} \equiv \hat{L}_{RC}\, g_m\left(\frac{m+1}{2m}[1 - V]s^2\right) \quad (8)$$

Here Vs^2 is the estimated variance of $\mathbf{X} \cdot \hat{\boldsymbol{\beta}}$, m is the number of observations used to calibrate the model minus the number of parameters in the model, and $g_m(\ )$ is a Bessel function [see *Cohn et al.*, 1989]. Cohn et al. use results from *Bradu and Mundlak* [1970] to derive the exact bias and variance of individual load estimates $\hat{L}_{RC,i}$, $\hat{L}_{QMLE,i}$ or $\hat{L}_{MVUE,i}$. However, *Thomas* [1988, p. 512] points out that unbiasing methods depend upon the validity of generally untested hypotheses:

> The performance of this correction procedure is sensitive to the assumptions, particularly to that of normality in the log space [*Duan*, 1983]. The assumption of normality is hard to justify considering how basic rating data are produced, and poor performance is to be expected. Therefore, blanket application of this bias correction procedure cannot be recommended.

The Accuracy of Annual or Monthly Load Estimates

Annual or monthly load estimates are usually computed as the sum of the estimated daily (or more frequent) loads. If we define the daily loads to be $\{\hat{L}_{*,1}, \cdots, \hat{L}_{*,N}\}$ for an arbitrary estimator (asterisk), and N is the total number of days over which we wish to compute the load, then

$$\hat{L}_{*,tot} = \sum_{i=1}^{N} \hat{L}_{*,i}. \quad (9)$$

Gilroy et al. [1990] derive the exact variance of $\hat{L}_{*,tot}$ for all of the estimators considered by *Cohn et al.* [1989], and also derive the variance of a seminonparametric smearing estimator [*Duan*, 1983].

Improvements to the Rating Curve: Additional Explanatory Variables

It has been noted that the rating curve model fails to capture some components of the structure of the concentration-discharge relationship. For example, *Miller* [1951, p. 11] finds that adjustments have to be made for different seasons:

TABLE 1. Study Sites and Descriptions

	Station Name			
	Choptank	Susquehanna	Patuxent	Potomac
Station ID	01491000	01578310	01594440	01646580
Drainage area, km^2	182	43600	560	18620
Mean annual flow, m^3/s	3.88	1190	11.5	353
Population density, people/km^2	761	222	1609	192
Percent agricultural	52	22	19	21
Percent urban	7	5	33	6
Nitrogen fertilizer, kg m^{-2}/yr^{-1}	0.0035	0.0007	0.0014	0.0011
Phosphorus fertilizer, kg m^{-2}/yr^{-1}	0.0024	0.0007	0.0010	0.0008
Storage, days of mean annual flow	0.0	2.2*	22.1†	0.0
Approximate number of samples	275	475	450	425

*Conowingo Reservoir, located immediately upstream from the sampling site for the Susquehanna River, has a capacity of 2.08×10^8 m^3.

†Howard Duckett Reservoir, 34 km upstream from sampling site, has a capacity of 2.12×10^7 m^3.

Separate curves were drawn for the spring run-off and the summer-winter-fall run-off because the plotted points defined different curves for these periods.

Glysson [1987] suggests stratifying the data according to the magnitude of flow, and applying a separate rating curve for each stratum. *Colby* [1956, p. 46] proposes that

> . . . more accurate sediment discharges are likely to be computed from a sediment rating curve that is shifted on the basis of occasional streamflow measurement and concurrent sediment samples.

In a data set comprising 4450 observations at 10-min intervals on the Mad River in California, Thomas observed that sediment concentrations were higher on the rising limb of the hydrograph than during the recession. *Glysson* [1987] concurs with this, although he also reports special cases where the opposite effect is seen. It is commonly recommended that different models be used for the two limbs of the hydrograph.

The Multivariate Rating Curve

An alternative to using multiple rating curves is to employ a single multivariate model. Time trends, seasonality, and other characteristics of the data, including hysteresis, can sometimes be described adequately by such a model. This approach usually requires fewer parameters than calibration of a separate model for each season, and therefore it requires less data to obtain a given degree of precision in load estimates. The bivariate models of (5), (7) and (8), when expressed in matrix notation, can be generalized to the multivariate case simply by changing the definition of $\mathbf{X}$ and $\boldsymbol{\beta}$.

Application of the Rating Curve to Constituents Other Than Sediment

The log linear models have been adapted for use with other constituents, in particular dissolved and sediment-related nutrients [*Verhoff et al.*, 1980; *Richards and Holloway*, 1987; *Young et al.*, 1988; *Preston et al.*, 1989; *Cohn et al.*, 1989; *Gilroy et al.*, 1990]. The models appear to have the same problems with nutrients that occur with sediment, although some of the effects are not as severe. For example, the hysteresis effect appears to be less pronounced for nutrients than it is for sediments. However, seasonality (differences between fall and spring, for example) may be much more important for nutrients than they are for sediment.

DATA SOURCES

The data considered in this study were collected between 1980 and 1988 at four sites representing the major tributaries to the Upper Chesapeake Bay: the Susquehanna, Potomac, Patuxent and Choptank rivers. Table 1 reports some of the characteristics of each of the sites and the upstream drainage basins. Three of the sampling sites are adjacent to U.S. Geological Survey (USGS) stream-gaging stations whose discharge records are rated "good" or "excellent." At the fourth site, which is on the Potomac River, water quality samples are collected at Chain Bridge, about 2 km downstream from the corresponding stream gage at Little Falls. However, the intervening drainage area between Little Falls and Chain Bridge accounts for only 16 km^2 in a basin of 18,600 km^2.

Three species of nitrogen (NH_4^+, $NO_2^- + NO_3^-$, and total nitrogen) and three species of phosphorus (PO_4, total dissolved P, and total P) were considered in this study. All samples were collected using USGS protocols [*Guy and Norman*, 1982] or equivalent methods. Samples were analyzed at one of four different laboratories. A quality control/quality assurance program is conducted to assure that results from the different laboratories are comparable (L. D. Zynjuk et al., manuscript in preparation, 1992).

A small number of samples, fewer than 20% in the worst case, were reported as "no detects." Each such "no detect" was replaced by its corresponding "detection limit." Because the number of such cases was small, it was determined that the use of this simple substitution method would not substantially bias the results (for discussion, see *Helsel and Cohn* [1988]).

Table 1 also reports the approximate number of samples available at each site. The numbers are approximate because some samples were not analyzed for all of the constituents, and thus the effective number of samples differs depending upon which constituent is considered.

PART 1: MODELING CONCENTRATIONS IN LOG SPACE

As the first part of this study, linear models were developed for each constituent at each site using all of the

observations available. A standard multivariate log linear model was employed in all cases. The model employs functions of two explanatory variables, flow and time, which are assumed (for now) to be measured continuously and without error.

In general, one has a set of N observations of a constituent concentration, denoted $\{C_1, \cdots, C_N\}$, along with the corresponding time of sampling $\{T_1, \cdots, T_N\}$ and the corresponding discharge $\{Q_1, \cdots, Q_N\}$ associated with each sample. For the purpose of developing the model, it is assumed that the logarithms of the concentrations obey a linear model of the form

$$\ln [C] = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{10}$$

where $\mathbf{X}$ is an $N \times k$ matrix, $\boldsymbol{\beta}$ is a $k \times 1$ matrix, and $\boldsymbol{\varepsilon}$ is an $N \times 1$ matrix of independent identically distributed normal random variables with mean zero and variance σ_ε^2.

A Seven-Parameter Log Linear Model for Constituent Concentrations

We had previously found that a particular linear model satisfactorily describes much of the variability in constituent concentrations. It requires estimation of a total of $k=7$ parameters: a constant; a quadratic fit to the logarithm of discharge (two parameters); a quadratic fit to time (two parameters); and a sinusoidal (first-order Fourier) function to remove the effects of annual seasonality (two parameters). The model can be written in the following form:

$$\ln [C] = \beta_0 + \beta_1 \ln [Q/\bar{Q}] + \beta_2\{\ln [Q/\bar{Q}]\}^2 + \beta_3[T - \bar{T}] + \beta_4[T - \bar{T}]^2 + \beta_5 \sin [2\pi T] + \beta_6 \cos [2\pi T] + \varepsilon \tag{11}$$

where ln[ ] denotes the natural logarithm function, Q is the discharge, and T is time measured in years. The errors, denoted ε, are assumed to be independent and normally distributed with zero mean and variance σ_ε^2. The β are parameters of the model which must be estimated from the data, and $\bar{Q}$ and $\bar{T}$ are centering variables.

The parameters related to seasonality, β_5 and β_6, can be converted into an amplitude and peak day, A and D^*, respectively, using the relationships

$$A = (\beta_5^2 + \beta_6^2)^{1/2} \tag{12}$$

$$D^* = \frac{365}{2\pi}\left[\tan^{-1}\left(\frac{\beta_5}{\beta_6}\right)\right] \tag{13}$$

Note that (13) will be either the maximum day or minimum day. If D^* corresponds to a minimum, the maximum occurs 6 months later.

Approximate standard errors of A and D^* can also be computed using first-order Taylor series:

$$SE[A] \approx \frac{[\beta_5^2 \text{ Var } (\beta_5) + \beta_6^2 \text{ Var } (\beta_6)]^{1/2}}{A} \tag{14}$$

$$SE[D^*] \approx \frac{365}{2\pi} \frac{[\beta_5^2 \text{ Var } (\beta_6) + \beta_6^2 \text{ Var } (\beta_5)]^{1/2}}{A^2} \tag{15}$$

where Var ( ) refers to the variance of the estimator, which is equal to the square of the standard errors reported in the regression.

$\bar{T}$ (and analogously $\bar{Q}$) is a "centering" variable that simplifies the numerical work and has no effect on the load estimates. It is defined so that the predictor variables corresponding to $\hat{\beta}_3$ and $\hat{\beta}_4$ are orthogonal. Specifically,

$$\bar{T} = \bar{T} + \frac{\sum_{i=1}^{N} (T_i - \bar{T})^3}{2\sum_{i=1}^{N} (T_i - \bar{T})^2} \tag{16}$$

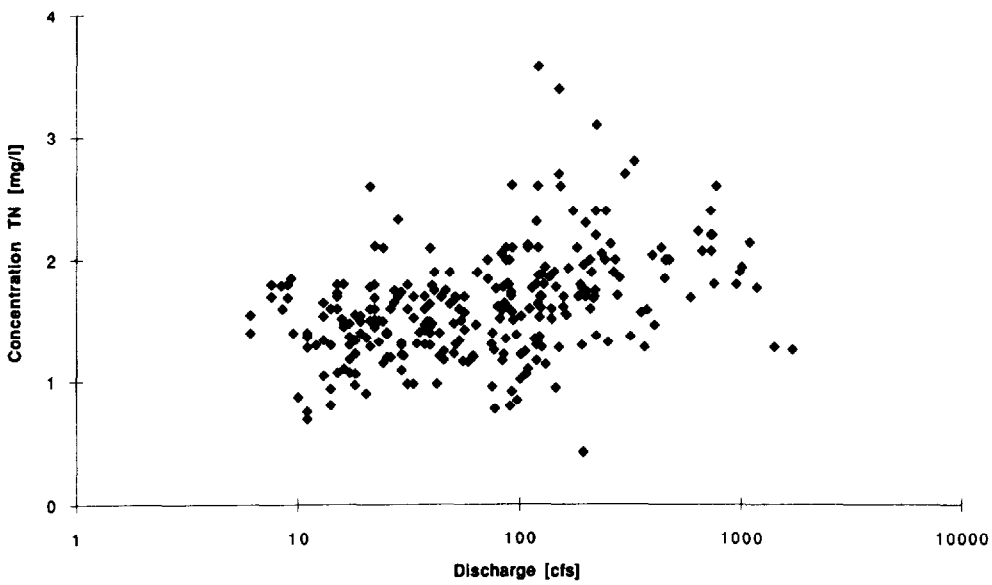

Fig. 1. Concentration of total nitrogen (TN) plotted against discharge for the Choptank River (station 01491000). 1 cubic foot per second (cfs) equals 2.8317×10^{-2} m^3/s.

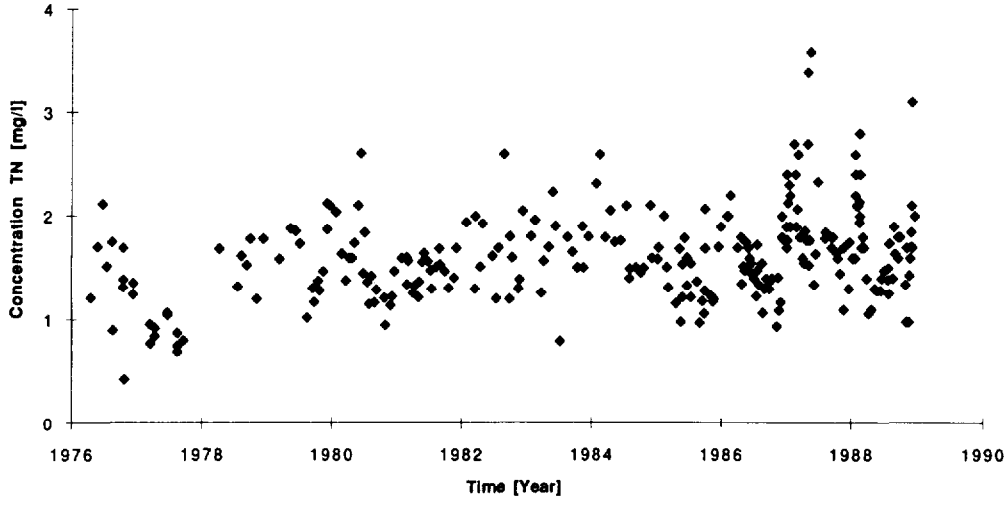

Fig. 2. Concentration of total nitrogen (TN) plotted against time for the Choptank River (station 01491000).

where

$$\bar{T} = \frac{1}{N} \sum_{i=1}^{N} T_i \tag{17}$$

The model described by (11) is attractive for several reasons. It is sufficiently flexible to capture many of the characteristics that have been observed in similar data; it is parsimonious, employing only seven parameters to deal with flow dependence, seasonality, and time trends; and it has generally been found to "explain" between 10 and 50% of the variability observed in the logarithms of constituent concentration data. However, the model is designed to characterize the data; there is not necessarily any explicit correspondence between the mathematical model and physical processes.

General Results From Fitting the Model With All Data

The model was fit to the six constituents at the four sites, yielding a total of 24 sets of estimates. In every case, the models showed statistically significant lack of fit (based on the MINITAB XLOF subcommand). The residuals always turned out to be significantly nonnormal according to the probability plot correlation coefficient (PPCC) test for normality [*Vogel*, 1986]. This was sometimes caused by the presence of one or two outliers, mostly on the low side. However, when the residuals were plotted against the explanatory variables and against the predicted values, they appeared to be reasonably homoscedastic. Aside from occasional outliers (which were not excluded from the data set), a visual inspection of residual plots and histograms did not suggest a substantial lack of fit. Figures 1–5 show the data,

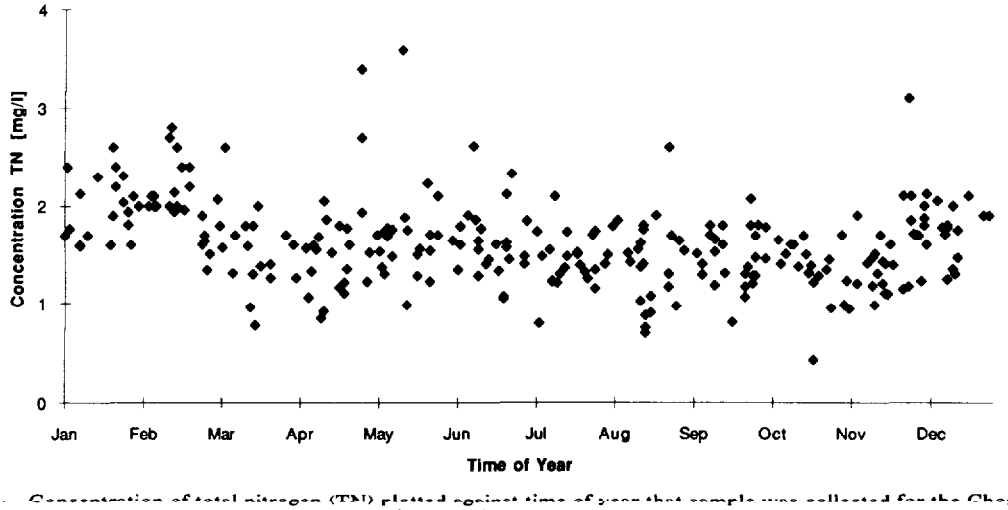

Fig. 3. Concentration of total nitrogen (TN) plotted against time of year that sample was collected for the Choptank River (station 01491000).

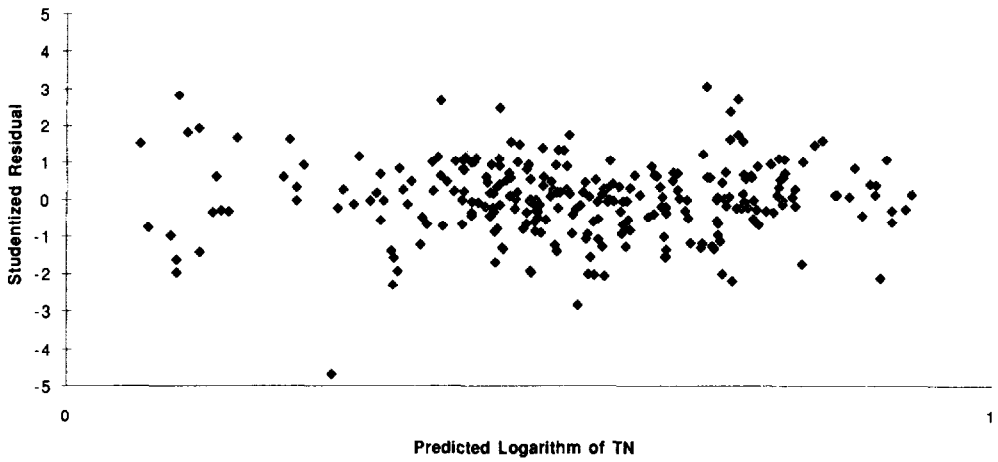

Fig. 4. Residuals plotted against predicted values from regression of the logarithm of total nitrogen (TN) on six predictor variables for the Choptank River (station 01491000).

the residuals and the predictions for a typical case.

Although the data were not collected uniformly in time, the serial correlation of the residuals was computed to provide some insight into whether the variance of the parameter estimates might be underestimated by ordinary least squares [see *Judge et al.*, 1980, p. 171]. The sample serial correlation was positive in 23 out of the 24 cases, and averaged approximately 0.25. The maximum value observed was 0.48.

Characteristics of the Fitted Models

Table 2 reports the values of the coefficients and their standard errors for each of the seven parameters at each of the four sites. Parameter estimates that are statistically significantly different from zero are noted with an asterisk. However, because there was no a priori reason to believe that the parameter values should be equal to zero, all explanatory variables were included in the model regardless of whether or not the corresponding parameter estimates were significantly different from zero.

It is worth considering briefly the parameter estimates. The parameter $\hat{\beta}_1$, which corresponds to the linear dependence on flow, differed significantly from zero for all six constituents at every site except at the Susquehanna. At the Susquehanna, $\hat{\beta}_1$ was significantly different from zero only for total dissolved phosphorus and for total phosphorus. This is perhaps explained by the presence of the Conowingo Reservoir immediately upstream from the Susquehanna sampling site. The quadratic flow term, $\hat{\beta}_2$, was significant in 13 cases. The parameters $\hat{\beta}_3$ and $\hat{\beta}_4$, which relate to time trends in the concentration, were significant in 16 cases, while seasonality, as measured by the presence of a significant amplitude, A, was statistically significant in 20 of the 24 cases.

Interpreting the Results

Many of the parameter values can be directly related to basin characteristics or to physical processes. For example, the value of β_1, which corresponds to the linear dependence of concentration on flow, may depend on the source of the constituent. Negative values indicate a dilution effect, suggesting point sources; near-zero values imply no flow effect, which can be expected for dissolved constituents where base flow and direct surface runoff have near-equal concentrations; and positive values suggest sediment-related, nonpoint sources.

In this study, negative coefficients were observed only for the Patuxent River. One would expect point sources to dominate in this basin, since this is the only urbanized watershed among the four studies. The nonpoint source loads are overwhelmed by the stable point source discharges. Thus the nutrient loads are relatively constant, and concentrations fall during high flows because of dilution. The standard deviation of the residuals, s, was also lower, for every constituent, at this site than at any of the others.

The value of β_3 corresponds to the magnitude of the log linear component of upward or downward time trends in concentration. It is interesting to note that there is a substantial (between 13 and 25% per year, depending on the particular species) downward trend in the phosphorus concentrations observed on the Patuxent River. This may be related to the implementation of a phosphorus control policy for the sewage treatment plants located upstream of the Patuxent monitoring site. Improved wastewater treatment, phased in between 1981

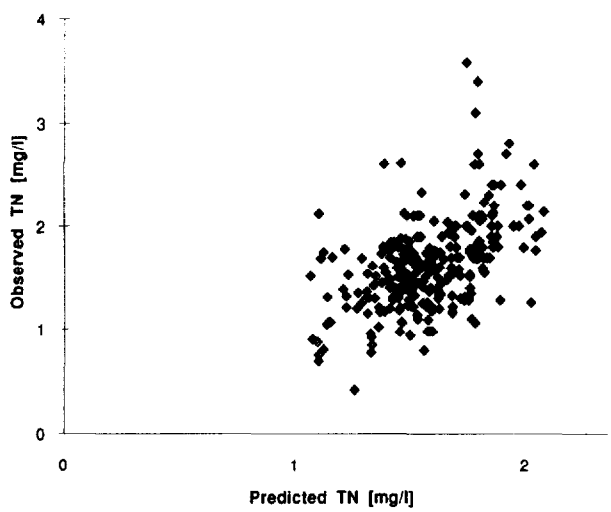

Fig. 5. Observed concentration of total nitrogen (TN) plotted against predicted values for the Choptank River (station 01491000).

TABLE 2*a*. Model Parameter Estimates Based on the Complete Data Sets: Nitrogen Species

	Choptank		Susquehanna		Patuxent		Potomac	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
				NH_4				
Constant	−2.823*	0.086	−2.399*	0.073	−0.726*	0.041	−3.199*	0.071
ln $[Q/Q_0]$	0.312*	0.055	−0.070	0.038	−0.681*	0.028	0.241*	0.035
ln $[Q/Q_0]^2$	−0.005	0.029	0.006	0.018	0.116*	0.024	0.108*	0.020
$T - T_0$	−0.019	0.013	0.021*	0.009	−0.041*	0.006	−0.003	0.017
$[T - T_0]^2$	−0.026*	0.004	−0.010*	0.003	−0.019*	0.002	−0.034*	0.007
Amplitude	0.235*	0.076	0.224*	0.060	0.549*	0.032	0.053	0.067
Peak day	July 30	75	March 3	61	March 11	18	March 2	270
Statistical summary								
s	0.77		0.74		0.45		0.78	
N	270		480		604		422	
R^2 for concentration	29.2		6.5		59.5		29.0	
R^2 for load	81.1		73.3		56.5		84.8	
				$NO_2 + NO_3$				
Constant	−0.019	0.031	0.094*	0.029	0.804*	0.024	0.291*	0.045
ln $[Q/Q_0]$	−0.137*	0.021	0.007	0.015	−0.456*	0.016	0.148*	0.022
ln $[Q/Q_0]^2$	−0.007	0.011	−0.006	0.007	0.059*	0.014	−0.121*	0.013
$T - T_0$	0.011*	0.005	0.015*	0.004	0.020*	0.004	0.007	0.010
$[T - T_0]^2$	0.003	0.002	−0.002	0.001	−0.006*	0.001	0.002	0.004
Amplitude	0.290*	0.036	0.289*	0.024	0.077*	0.017	0.358*	0.039
Peak day	March 16	32	Feb. 28	30	Aug. 7	101	Feb. 2	33
Statistical summary								
s	0.29		0.31		0.24		0.50	
N	279		506		428		432	
R^2 for concentration	24.9		32.3		74.8		44.6	
R^2 for load	94.4		94.9		77.4		92.3	
				Total Nitrogen				
Constant	0.464*	0.026	0.549*	0.023	1.281*	0.020	0.743*	0.023
ln $[Q/Q_0]$	0.056*	0.017	0.022	0.012	−0.345*	0.014	0.251*	0.012
ln $[Q/Q_0]^2$	0.002	0.009	0.014*	0.006	0.092*	0.012	−0.011	0.007
$T - T_0$	0.023*	0.004	0.017*	0.003	0.014*	0.004	0.025*	0.005
$[T - T_0]^2$	−0.003*	0.001	−0.005*	0.001	−0.009*	0.001	−0.004	0.002
Amplitude	0.055	0.029	0.154*	0.019	0.016	0.015	0.144*	0.019
Peak day	March 28	151	Feb. 25	51	Aug. 3	450	Dec. 28	61
Statistical summary								
s	0.24		0.25		0.20		0.26	
N	274		499		418		426	
R^2 for concentration	26.3		28.0		70.8		66.5	
R^2 for load	96.8		96.5		87.6		97.9	

Here, s denotes the standard deviation of the residuals from ordinary least squares fit; N, the number of observations used to fit the model; and R^2 for concentration and load, the "variability explained" by the models for logarithm of concentration and load, respectively.
*Statistically different from zero at 5% level.

and 1989, in combination with imposition of Maryland's phosphate detergent ban in December 1985, has resulted in more than a 50% reduction in point source phosphorus loadings to the river since 1981. Substantial downtrends were also observed for PO_4 on the Choptank and Susquehanna rivers. However, most of the nitrogen constituent trends were upward with a magnitude of between 1 and 3% per year.

PART 2: ESTIMATING THE LOADS

In the first part of this study, models were developed for estimating the nutrient concentration in "log" space. The purpose of this section is to investigate the validity of load estimators based on such log linear models.

The approach here is similar to that of *Thomas* [1988], and involves repeated split-sampling studies. Although between 300 and 500 observations were, in fact, collected at each of the four sites, one can imagine asking how well one would have estimated the total load given only 75 observations, the approximate number available at most National Stream Quality Accounting Network (NASQAN) sites around the United States.

The following experiment suggests itself: One could randomly select 75 observations from the total sample for each of the 24 data sets, use these 75 to calibrate a concentration model, and use the model to estimate the sum of the instantaneous loads for the, for example, 500 moments in time when there were measurements. These sums are denoted $\hat{L}_{RC,tot}$, $\hat{L}_{QMLE,tot}$, and $\hat{L}_{MVUE,tot}$ for the three estimators. Each of these sums can be compared to the sum of the observed loads, denoted $L_{OBS,tot}$. One expects the two values, on average, to be close to each other for an unbiased estimator. The experiment is then repeated by selecting different random samples. Repeated subsampling provides an indication of the distribution of estimates, as well as a better estimate of the bias of each approach.

TABLE 2b. Model Parameter Estimates Based on the Complete Data Sets: Phosphorus Species

	Choptank		Susquehanna		Patuxent		Potomac	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
				PO_4				
Constant	−3.328*	0.098	−4.152*	0.082	−1.313*	0.044	−3.201*	0.056
ln $[Q/Q_0]$	0.273*	0.058	0.015	0.041	−0.563*	0.030	0.213*	0.028
ln $[Q/Q_0]^2$	−0.019	0.030	−0.066*	0.019	0.083*	0.026	−0.039*	0.016
$T - T_0$	−0.118*	0.014	−0.122*	0.010	−0.212*	0.006	0.024	0.014
$[T - T_0]^2$	−0.014*	0.004	−0.014*	0.004	−0.032*	0.002	−0.036*	0.005
Amplitude	0.684*	0.100	0.168*	0.060	0.277*	0.035	0.429*	0.051
Peak day	Sept. 3	24	Feb. 5	87	Oct. 17	36	Nov. 18	31
Statistical summary								
s	0.78		0.80		0.48		0.64	
N	244		483		588		430	
R^2 for concentration	41.4		29.8		82.8		30.3	
R^2 for load	77.2		76.7		75.2		87.2	
				Total Dissolved Phosphorus				
Constant	−3.717*	0.105	−4.427*	0.096	−1.464*	0.057	−2.897*	0.053
ln $[Q/Q_0]$	0.174*	0.056	0.075*	0.038	−0.484*	0.032	0.290*	0.031
ln $[Q/Q_0]^2$	0.048	0.031	0.055*	0.018	0.009	0.029	−0.013	0.016
$T - T_0$	−0.011	0.015	0.019	0.010	−0.197*	0.006	0.042*	0.012
$[T - T_0]^2$	−0.009	0.006	0.015*	0.006	−0.012*	0.003	−0.029*	0.004
Amplitude	0.606*	0.098	0.086	0.058	0.321*	0.038	0.462*	0.050
Peak day	Sept. 15	26	Dec. 8	168	Oct. 16	32	Nov. 10	27
Statistical summary								
s	0.63		0.68		0.46		0.57	
N	165		383		473		401	
R^2 for concentration	26.8		5.2		82.6		33.2	
R^2 for load	79.0		79.0		76.2		88.5	
				Total Phosphorus				
Constant	−2.716*	0.060	−3.091*	0.058	−0.664*	0.037	−2.129*	0.053
ln $[Q/Q_0]$	0.376*	0.037	0.228*	0.030	−0.159*	0.025	0.578*	0.026
ln $[Q/Q_0]^2$	0.000	0.017	0.113*	0.014	0.082*	0.022	0.074*	0.015
$T - T_0$	−0.016*	0.008	−0.013	0.007	−0.127*	0.005	0.002	0.009
$[T - T_0]^2$	0.000	0.003	0.000	0.003	−0.018*	0.002	−0.011*	0.003
Amplitude	0.498*	0.060	0.133*	0.047	0.276*	0.029	0.561*	0.052
Peak day	Sept. 1	26	Sept. 1	90	Oct. 25	33	Oct. 4	21
Statistical summary								
s	0.57		0.60		0.42		0.62	
N	314		504		634		473	
R^2 for concentration	29.5		22.5		67.0		60.1	
R^2 for load	87.4		85.5		75.3		92.0	

Here, s denotes the standard deviation of the residuals from ordinary least squares fit; N, the number of observations used to fit the model; and R^2 for concentration and load, the "variability explained" by the models for logarithm of concentration and load, respectively.
*Statistically different from zero at 5% level.

Experimental Design

For each of the 24 data sets, 100 subsamples of 75 observations were randomly drawn from the N (ranging from 165 to 634) observations available. Each subsample was drawn without replacement, simulating the situation that might have occurred had only 75 samples been collected. The model described by (11) was then fit to each of the 100 subsamples, and, using the methods described by *Cohn et al.* [1989], estimates $\hat{L}_{RC,tot}$, $\hat{L}_{QMLEC,tot}$, and $\hat{L}_{MVUE,tot}$ were obtained for the sum of the N instantaneous loads. Additionally, the predicted variance of $\hat{L}_{MVUE,tot}$ was computed for one fourth of the subsamples (this step requires large amounts of computer time, and sufficient precision was available from 25 subsamples), using the methods described by *Gilroy et al.* [1990].

Finite Population Correction for the Variance

The variances given by Gilroy et al. are not quite right for the case here, because they assume that each successive subsample of 75 is drawn anew from an infinite population, while the sampling design here involves repeated sampling from a finite population. This will result in lower variances among the observed subsamples than would be predicted by the equations of Gilroy et al., because the subsamples will share observations [see *Stuart*, 1976].

A standard finite population correction [*Stuart*, 1976] exists for estimating the variance of the mean, $\overline{X_K} \equiv \sum X_i/K$, of a random variable based on repeated subsamples of size K drawn from a finite sample of size N:

$$\text{var}(\overline{X_K}) = \text{var}(X)\left\{\frac{N-K}{(N-1)N}\right\} \qquad (18)$$

TABLE 3. Observed and Theoretical Correction Factors When Sampling From a Finite Population Based on 2500 Monte Carlo Experiments (50 Subsamples Drawn From 50 Finite Populations; Subsamples of 75 Selected Without Replacement)

	Choptank	Susquehanna	Patuxent	Potomac
NH_4				
Observed R_1	0.74	0.86	0.88	0.80
Theoretical R_1	0.72	0.85	0.88	0.82
$NO_2 + NO_3$				
Observed R_1	0.77	0.88	0.84	0.91
Theoretical R_1	0.73	0.85	0.83	0.83
Total nitrogen				
Observed R_1	0.77	0.85	0.84	0.86
Theoretical R_1	0.73	0.85	0.82	0.83
PO_4				
Observed R_1	0.80	0.86	0.87	0.85
Theoretical R_1	0.70	0.85	0.87	0.83
Total dissolved phosphorus				
Observed R_1	0.65	0.81	0.86	0.81
Theoretical R_1	0.55	0.81	0.84	0.82
Total phosphorus				
Observed R_1	0.83	0.86	0.88	0.90
Theoretical R_1	0.76	0.85	0.88	0.84

The case here is somewhat different, since we are interested in the variance of a function of repeated subsamples, the sum of the exponentiated predictions based on the fitted regression model, which relies on all of the data. This function could assume $\binom{N}{K}$ distinct values.

Equation (18) is asymptotically correct in certain limiting conditions. However, no general result is available. To determine whether the correction factor is appropriate for the situation encountered here, a set of Monte Carlo sampling experiments was run to simulate the experiment that was conducted with the real data. For each of the four rivers and six constituents, the parameter values $\{\hat{\beta}, \hat{\sigma}^2\}$ were computed. Then 50 new data sets were generated, using the equation $\bar{\mathbf{C}}_i \equiv \mathbf{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{\varepsilon}$ is a vector of independent random normal $N(0, \hat{\sigma}^2)$ variates. Each $\bar{\mathbf{C}}_i$ had the same number of observations as the original complete data set. Then the sampling experiment that was done with the real data was conducted using these simulated "finite populations," and the statistical behavior of the estimates was recorded.

Fifty subsamples of size 75 were then drawn (without replacement for each sample) from each $\bar{\mathbf{C}}_i$. The seven-parameter model was fit to these samples, and the sum of the loads corresponding to $\mathbf{X}$, denoted $\bar{L}_{i,j}$, was calculated. The variance of the 50 estimated loads corresponding to each $\bar{\mathbf{C}}_i$, denoted $\bar{V}_j$, was calculated:

$$\bar{V}_i \equiv \text{Var}\,(\bar{L}_{i,*}) = \frac{1}{50}\sum_{i=1}^{50}\left\{\frac{1}{49}\sum_{j=1}^{50}[\bar{L}_{i,j} - \text{Mean}\,(\bar{L}_{i,*})]^2\right\} \tag{19}$$

where

$$\text{Mean}\,(\bar{L}_{i,*}) = \frac{1}{50}\sum_{j=1}^{50}\bar{L}_{i,j} \tag{20}$$

Similarly, the variance of the infinite population was estimated by computing $\bar{V}_j$:

$$\bar{V}_j \equiv \text{Var}\,(\bar{L}_{*,j}) = \frac{1}{50}\sum_{j=1}^{50}\left\{\frac{1}{49}\sum_{i=1}^{50}[\bar{L}_{i,j} - \text{Mean}\,(\bar{L}_{*,j})]^2\right\} \tag{21}$$

Although the terms within the braces are not independent, $\bar{V}_j$ does provide an unbiased estimate of the variance of the theoretical "infinite population." The ratio of $\bar{V}_i$ to $\bar{V}_j$, denoted R_1, provides an estimate of the effect of sampling from a finite population. Table 3 contains estimates of R_1 as well as the theoretical correction formula computed using (18).

It was found that the Monte Carlo results were not, in general, statistically significantly different from the results given by (18). Apparently the large-sample results are satisfactory for the cases that arose here, although one can imagine situations where this would not be true. Thus the "finite population" correction factor was used to adjust the predicted variances for the case of an infinite population (given by *Gilroy et al.* [1990]).

It should be noted that the use of simulation, as was done here, obviates the need for the analytical results of Gilroy et al. In most cases, however, the analytical results would be substantially easier to employ. In this case we were forced to use simulation because of the particular sampling scheme we employed.

Table 4 reports the average observed coefficient of variation (CV) along with the average predicted CV for $\hat{L}_{\text{MVUE,tot}}$, where the latter is adjusted by (18) to account for sampling

TABLE 4. Predicted and Observed Percent Coefficient of Variation of Load Estimates Based on Subsamples of 75 Selected Without Replacement (Predictions Adjusted to Account for Sampling From a Finite Population)

	Choptank	Susquehanna	Patuxent	Potomac
NH_4				
Observed	17	10	6	18
Predicted	20	12	7	26
Standard deviation	3	1	1	3
t statistic	0.9	1.0	1.0	3.1
$NO_2 + NO_3$				
Observed	6	4	3	7
Predicted	5	5	3	11
Standard deviation	1	1	1	1
t statistic	−1.0	0.9	−0.7	3.6
Total nitrogen				
Observed	17	13	6	27
Predicted	19	13	7	20
Standard deviation	3	2	1	4
t statistic	0.9	0.2	1.4	−1.8
PO_4				
Observed	12	10	5	16
Predicted	13	11	6	16
Standard deviation	2	1	1	2
t statistic	0.9	0.8	1.4	0.0
Total dissolved phosphorus				
Observed	5	4	4	11
Predicted	5	4	3	7
Standard deviation	1	1	1	2
t statistic	0.3	0.4	−1.5	−2.4
Total phosphorus				
Observed	13	10	9	22
Predicted	15	11	8	25
Standard deviation	2	1	1	3
t statistic	1.4	1.3	1.2	0.7

TABLE 5. Percent Difference Between Average Estimated Loads Based on Subsamples and Load Computed From Complete Data (Based on Subsamples of 75 Selected Without Replacement)

	Choptank			Susquehanna			Patuxent			Potomac		
	$\hat{L}_{RC}$	$\hat{L}_{QMLE}$	$\hat{L}_{MVUE}$	$\hat{L}_{RC}$	$\hat{L}_{QMLE}$	$\hat{L}_{MVUE}$	$\hat{L}_{RC}$	$\hat{L}_{QMLE}$	$\hat{L}_{MVUE}$	$\hat{L}_{RC}$	$\hat{L}_{QMLE}$	$\hat{L}_{MVUE}$
NH_4	−19	10	4	−22	3	1	−7	2	1	−3	32	25
standard deviation	16	16	16	7	7	7	3	3	3	14	14	14
t statistic	−1.2	0.6	0.3	−3.2	0.4	0.1	−2.7	0.9	0.4	−0.2	2.3	1.8
$NO_2 + NO_3$	−2	2	1	0	4	4	−3	0	0	−2	12	10
standard deviation	3	3	3	2	2	2	1	1	1	5	5	5
t statistic	−0.6	0.6	0.5	−0.2	2.1	2.0	−1.7	0.2	0.0	−0.4	2.5	2.1
Total nitrogen	−1	2	2	−2	1	1	−2	0	0	−10	−7	−8
standard deviation	3	3	3	2	2	2	1	1	1	4	4	4
t statistic	−0.2	0.8	0.7	−1.3	0.5	0.4	−1.3	0.1	−0.1	−2.4	−1.7	−1.8
PO_4	−12	19	13	−24	6	3	−7	4	3	7	32	26
standard deviation	14	14	14	9	9	9	3	3	3	9	10	9
t statistic	−0.9	1.4	0.9	−2.7	0.7	0.4	−2.5	1.6	1.1	0.8	3.3	2.8
Total dissolved phosphorus	−5	16	12	−14	9	7	−6	5	4	5	24	20
standard deviation	12	12	12	6	6	6	3	3	3	8	8	8
t statistic	−0.4	1.3	1.0	−2.3	1.5	1.2	−1.9	1.6	1.3	0.6	3.0	2.4
Total phosphorus	−11	4	1	−21	−5	−6	−5	3	2	−10	9	5
standard deviation	9	9	9	6	6	6	3	3	3	15	15	15
t statistic	−1.3	0.5	0.1	−3.3	−0.8	−1.0	−1.6	1.1	0.6	−0.7	0.6	0.3

from a finite population. Aside from a few outliers (which corresponded to data sets with outliers), the predicted CVs were not significantly different from the observed CVs; this is apparent in the t statistics. As expected, the estimated variances did not appear to be biased either upward or downward.

Statistical Aspects of Testing for Bias in the Mean

It was also desired to test whether the estimated loads based on a particular estimator (say, $\hat{L}_{*,tot}$, where the asterisk corresponds to MVUE, RC or QMLE), were significantly different from the true load, denoted L_{tot}. A standard statistical approach would be to assume that $\hat{L}_{*,tot}$, is approximately normally distributed, and then to test whether the average of the 100 estimates, $\langle \hat{L}_{*,tot} \rangle$ is close to L_{tot}. Unfortunately, we cannot measure L_{tot} directly. However, under the null hypothesis, $L_{OBS,tot}$ is normally distributed with mean L_{tot}. So one can construct an approximate t test (the number of degrees of freedom is assumed to be large) by comparing the two random variables $\hat{L}_{*,tot*}$ to $L_{OBS,tot}$:

$$T = \frac{L_{OBS,tot} - \langle \hat{L}_{*,tot} \rangle}{[\text{Var}\,(L_{OBS,tot}) + \text{Var}\,(\langle \hat{L}_{*,tot} \rangle)]^{1/2}} \qquad (22)$$

The variance of $L_{OBS,tot}$ can be computed [*Aitchison and Brown*, 1957] by assuming that the errors are independent lognormal variates, and thus

$$\text{Var}\,(L_{OBS,tot}) = [\exp\,(\hat{\sigma}_\varepsilon^2) - 1] \sum_{i=1}^{N} L_{OBS,i}^2 \qquad (23)$$

The variance of $\langle \hat{L}_{*,tot} \rangle$ is available from *Gilroy et al.* [1990], after making the correction for subsampling a finite population, or it can be computed from the 100 replicates.

Table 5 reports the bias results as a 3 × 3 matrix for each of the 24 cases. Each column corresponds to one of the estimators, $\hat{L}_{RC,tot}$, $\hat{L}_{QMLE,tot}$, and $\hat{L}_{MVUE,tot}$, respectively. The first row reports the mean percentage bias as:

$$B \equiv 100 \left[\frac{\hat{L}_{*,tot*} - L_{OBS,tot}}{L_{OBS,tot}} \right] \qquad (24)$$

the second row is the estimated standard error of B, and the third row is the corresponding approximate t statistic.

The greatest upward deviation, 32%, occurred for $\hat{L}_{QMLE,tot}$ when estimating PO_4 and NH_4 for the Potomac River. The average of the t statistics corresponding to $\hat{L}_{RC,tot}$ was significantly less than zero, indicating a negative bias in the estimator, and the average t statistic corresponding to $\hat{L}_{QMLE,tot}$ was significantly biased upward. The average bias of the t statistics corresponding to the $\hat{L}_{MVUE,tot}$ was not statistically significant. The PPCC test was used to test the sample of 24 t statistics for normality. When sample mean and variances were used, rather than the assumed values of zero and one, the assumption of normality could not be rejected at the 5% level for any of the estimators.

DISCUSSION

Recent research on the rating curve, and log linear models in general, has suggested that there may be substantial dangers in blindly applying log linear model theory to water quality data. For example, *Thomas* [1988] has shown that, with respect to sediment loads in a small watershed, the standard assumption of log linearity can be invalid. However, the experience here, which involves nutrient transport from relatively large watersheds, suggests that log linear models may be satisfactory for some purposes. Many of Thomas's concerns are warranted; for example, the log linear model employed here exhibited statistically significant lack of fit to the nutrient data. However, apparently the statistical properties of the load estimators, in particular the MVUE, were not seriously affected by the model's lack of fit. Load estimates based on samples of size 75, a relatively small number, were found to have sampling properties predicted well by log linear model theory. For example, for total nitrogen (TN) the average percent difference between the estimated loads and the observed loads ranged from

−8% to +2% at the four sites, while the corresponding differences for total phosphorus (TP) were −6% to +5%. These differences are relatively unsubstantial, and none was statistically significant. Moreover, the observed CVs for TN and TP, which ranged from 7% to 25%, were predicted accurately by the statistical theory.

CONCLUSIONS

The validity of several fluvial load estimation methods was tested using repeated split-sample studies with nutrient data from some of the major tributaries to Chesapeake Bay. Several observations can be made:

1. Simple log linear models provided a useful and reasonably accurate description of nutrient concentrations in the streams examined here. However, statistically significant, though not substantial, lack of fit was observed in all cases.

2. Load estimates assuming the validity of a log linear relationship appear to be fairly insensitive to modest amounts of model misspecification or nonnormality of residual errors.

3. The variances of annual or monthly load estimates based on infrequent sampling appear to be well-described by linear model theory.

In summary, the MVUE load estimator employing log linear models was found to provide satisfactory estimates both of nutrient loads and of the uncertainties in total load estimates.

REFERENCES

Aitchison, J., and J. A. C. Brown, *The Lognormal Distribution With Special Reference to Its Uses in Economics*, Cambridge University Press, New York, 1957.

Bradu, D., and Y. Mundlak, Estimation in lognormal linear models, *J. Am. Stat. Assoc.*, *65*(329), 198–211, 1970.

Brown, H. E., E. A. Hansen, and N. E. Champagne, Jr., A system for measuring total sediment yield from small watersheds, *Water Resour. Res.*, *6*(3), 818–826, 1970.

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

Colby, B. R., The relationship of sediment discharge to streamflow, open file report, U.S. Geol. Surv., Reston, Va., 170 pp., 1956.

Draper, N. R., and H. Smith, *Applied Regression Analysis*, John Wiley, New York, 1981.

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

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

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

Glysson, G. D., Sediment transport curves, *U.S. Geol. Surv. Open File Rep.*, *87-218*, 47 pp., 1987.

Guy, H. P., and V. W. Norman, Field methods for measurement of fluvial sediment, *Techniques of Water Resources Investigations*, chap. C2, 59 pp., U.S. Geological Survey, Reston, Va., 1982.

Helsel, D. H., and T. Cohn, Estimation of descriptive statistics for multiply censored water quality data, *Water Resour. Res.*, *24*(12), 1997–2004, 1988.

Judge, G. E., W. E. Griffiths, R. C. Hill, and T-C. Lee, *The Theory and Practice of Econometrics*, John Wiley, New York, 1980.

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

Miller, C. R., Analysis of flow-duration, sediment-rating curve method of computing sediment yield, report, 15 pp., U.S. Bur. of Reclam., Denver, Colo., 1951.

Preston, S. D., V. J. Bierman, Jr., and S. E. Silliman, An evaluation of methods for the estimation of tributary mass loads, *Water Resour. Res.*, *25*(6), 1379–1389, 1989.

Richards, R. P., and J. Holloway, Monte Carlo studies of sampling strategies for estimating tributary loads, *Water Resour. Res.*, *23*(10), 1939–1948, 1987.

Stuart, A., *Basic Ideas of Scientific Sampling, Griffin's Stat. Monogr. and Courses*, vol. 4, 2nd ed., Hafner, New York, 1976.

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

Thomas, R. B., Monitoring baseline suspended sediment in forested basins: The effects of sampling of suspended sediment rating curves, *Hydrol. Sci.*, *33*(5), 499–514, 1988.

Thompson, M. P., H. Joe, and M. Church, Statistical modelling of sediment concentration, *Rep. IWD-HQ-SS-88-1*, 60 pp., Sediment Surv. Sect., Environ. Can., Ottawa, 1987.

Verhoff, F. H., S. M. Yaksich, and D. A. Melfi, River nutrient and chemical transport estimation, *J. Environ. Eng. Div. Am. Soc. Civ. Eng.*, *106*, 591–608, 1980.

Vogel, R. M., The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypotheses, *Water Resour. Res.*, *22*(4), 587–590, 1986.

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

Young, T. C., J. V. DePinto, and T. M. Heidtke, Factors affecting the efficiency of some estimators of fluvial total phosphorus load, *Water Resour. Res.*, *24*(9), 1535–1543, 1988.

D. L. Caulder, T. A. Cohn, and E. J. Gilroy, U.S. Geological Survey, MS 410, Reston, VA, 22092.

R. M. Summers, Maryland Department of the Environment, Special Projects, 2500 Broening Highway, Baltimore, MD, 21224.

L. D. Zynjuk, U.S. Geological Survey, 208 Carroll Building, 8600 LaSalle Road, Towson, MD, 21204.

Received February 5, 1992;
revised April 16, 1992;
accepted April 29, 1992.)