# Estimating contaminant loads in rivers: An application of adjusted maximum likelihood to type 1 censored data

Timothy A. Cohn
U.S. Geological Survey, Reston, Virginia, USA

Received 22 November 2004; revised 14 March 2005; accepted 23 March 2005; published 6 July 2005.

[1] This paper presents an adjusted maximum likelihood estimator (AMLE) that can be used to estimate fluvial transport of contaminants, like phosphorus, that are subject to censoring because of analytical detection limits. The AMLE is a generalization of the widely accepted minimum variance unbiased estimator (MVUE), and Monte Carlo experiments confirm that it shares essentially all of the MVUE's desirable properties, including high efficiency and negligible bias. In particular, the AMLE exhibits substantially less bias than alternative censored-data estimators such as the MLE (Tobit) or the MLE followed by a jackknife. As with the MLE and the MVUE the AMLE comes close to achieving the theoretical Frechet-Cramér-Rao bounds on its variance. This paper also presents a statistical framework, applicable to both censored and complete data, for understanding and estimating the components of uncertainty associated with load estimates. This can serve to lower the cost and improve the efficiency of both traditional and real-time water quality monitoring.

**Citation:** Cohn, T. A. (2005), Estimating contaminant loads in rivers: An application of adjusted maximum likelihood to type 1 censored data, *Water Resour. Res.*, *41*, W07003, doi:10.1029/2004WR003833.

## 1. Issue of "Load Estimation"

[2] Scientists and regulators increasingly recognize that monitoring fluvial transport of contaminants is a useful tool for watershed science and management (see *National Research Council* [2001], *Government Accountability Office* [2003], and *Cohn* [1995] for discussion and further citations). However, direct measurement of contaminant transport is not, in general, easy to do. The instantaneous "load" passing through a stream cross section is the product of water discharge and constituent concentration. While continuously measuring the former is routine, measurement of contaminant concentrations typically demands careful collection and preservation of samples, followed by chemical analyses, which altogether can cost hundreds of dollars per observation. A further complication sometimes arises: laboratory analyses are subject to censoring because of "analytical detection limits," resulting in values reported as "less than" a detection level.

[3] This paper presents an efficient and nearly unbiased load estimator for use with censored data. The estimator is based on a generalization of the minimum variance unbiased estimator (MVUE) that has been employed with complete samples [*DeLong*, 1982; *Cohn et al.*, 1989, 1992a; *Gilroy et al.*, 1990a; *Crawford*, 1991, 1996]. The proposed estimator nearly achieves Frechet-Cramér-Rao (FCR) bounds in small samples, which define the theoretical minimum variance of an unbiased estimator, while exhibiting vanishingly little bias.

[4] Minimizing estimator bias is important because load estimates are often added together to compute overall transport from multiple tributaries into a single water body, and bias, to a greater extent than random variability, tends to accumulate. In addition, although scientists understand that "the term 'unbiased' should not be allowed to convey overtones of a non-technical nature" [*Stuart and Ord*, 1979, section 17.9], many people remain prejudiced against "bias." It is typically easier to avoid bias than to justify it to regulators, managers, or the public.

[5] This paper also presents a statistical framework, applicable to both censored and complete data, for understanding and estimating the components of uncertainty associated with load estimates. This can serve to improve the efficiency of monitoring networks, as well as to characterize the uncertainty in both traditional and real-time water quality monitoring systems.

## 2. Framing the Statistical Problem

[6] The "constituent load," denoted $\mathbf{L}$, is defined as the quantity of material transported through a specified river cross section during the time interval $\{t_a, t_b\}$. It can be expressed formally in terms of an integral:

$$\mathbf{L} \equiv \int_{t_a}^{t_b} L(t)dt = K_u \int_{t_a}^{t_b} C(t) \cdot Q(t)dt \quad (1)$$

where $L(t)$ is the "instantaneous load" (e.g., in kg/d), $K_u$ is a constant whose value depends on the units, and $C(t)$ and $Q(t)$ are, respectively, the concentration (e.g., in mg/L) and water discharge (e.g., in cubic feet/s) at time $t$.

[7] There is assumed to be a stream gauge that provides a continuous and accurate record of $Q(t)$ at the site.

However, there is no corresponding *continuous* record of $C(t)$. Instead, $C(t)$ is available only for a small number, $N$, of discrete times $\{t_1,\ldots, t_N\}$. In addition, some of the concentration data may be reported as less than a specified analytical detection limit. For example, when sample orthophosphate concentrations are below 0.02 mg/L, many laboratories will report a value of "<0.02 mg/L" rather than a single point estimate. Although the precise meaning of less than varies according to laboratory practice [*Childress et al.*, 1999], one can assume that "<0.02 mg/L" implies that the true value of the concentration is below 0.02 mg/L. Data described as less than a threshold value are referred to as "type 1 censored samples" [*David*, 1981].

### 2.1. Rating Curve and Related Models

[8] "Rating curves" have long been used to describe the relation between the logarithm of concentration and the logarithm of discharge [*Miller*, 1951; *Colby*, 1955; *Cohn*, 1995]. Fitted to the $N$ observed values $\vec{C} \equiv \{C_1,\ldots, C_N\}$, the rating curve permits estimation of $C(t)$ from $Q(t)$, and therefore allows evaluation of the integral in equation (1).

[9] The general form of the rating curve is often expressed as a linear model:

$$Y^\star(t) \equiv \ln(C(t)) = \mathbf{X}(t)\boldsymbol{\beta}^\star + \epsilon(t) \tag{2}$$

where $Y^\star(t) \equiv \ln(C(t))$, $\mathbf{X}(t) = \{1, \ln(Q(t)),\ldots X_K(t)\} = \{X_0(t), X_1(t),\ldots X_K(t)\}$ is a vector of predictor variables, and $\boldsymbol{\beta}^\star = \{\beta_0^\star, \beta_1^\star,\ldots \beta_K^\star\}'$ is a vector of $K+1$ regression coefficients ("real-time" monitoring programs often employ additional predictors such as turbidity and temperature to improve real-time water quality estimates [*Christensen and Ziegler*, 2000]). The linear model defined by equation (2) has been studied extensively and its properties, both statistical and hydrological, are well understood (see *Draper and Smith* [1981] and *Cohn* [1995] for further discussion and additional references).

[10] Substituting equation (2) into equation (1) yields

$$\begin{aligned}\mathbf{L} &= K_u \int_{t_a}^{t_b} C(t)\cdot Q(t)dt \\ &= K_u \int_{t_a}^{t_b} \exp(\mathbf{X}(t)\boldsymbol{\beta}^\star + \ln(Q(t)))\cdot \exp(\epsilon(t))dt\end{aligned} \tag{3}$$

The integral in equation (3) depends only on continuously measured variables (like discharge), fixed coefficients $\boldsymbol{\beta}^\star$, and the error $\epsilon(t)$.

[11] If, as assumed here, $\ln(Q)$ is used as a predictor variable in equation (2), the model described by equation (2) can be further simplified by moving the $\ln(Q(t))$ and $K_u$ into the regression to yield

$$\ln(L(t)) = \mathbf{X}(t)\boldsymbol{\beta} + \epsilon(t) \tag{4}$$

and finally

$$\mathbf{L} = \int_{t_a}^{t_b} \exp(\mathbf{X}(t)\boldsymbol{\beta})\cdot \exp(\epsilon(t))dt \tag{5}$$

Equations (3) and (5) represent exactly the same model [*Gilroy et al.*, 1990b] because

$$\boldsymbol{\beta} = \boldsymbol{\beta}^\star + \begin{bmatrix}\ln(K_u)\\ 1\\ 0\\ \ldots\\ 0\end{bmatrix} \tag{6}$$

### 2.2. Error Term, $\epsilon(t)$

[12] The "error," $\epsilon(t)$, represents the difference between the logarithm of the observed concentration and the value based on the linear model. Although in practice the error term includes measurement error and model error, we will assume here that the error consists primarily of natural variability and that measurement error can be neglected. It will be further assumed that $\epsilon(t)$ is normally distributed $N(0, \sigma^2)$, and that the serial correlation between $\epsilon_i$ and $\epsilon_j$ is negligible for $i \neq j$, or, more generally, that $\epsilon(t)$ is independent of $\epsilon(t+\delta)$ for values of $\delta$ on the order of the elapsed time between concentration measurements (in many cases about 2 weeks).

### 2.3. Parameter Estimation With Complete Data

[13] Given a vector of uncensored observations

$$\mathbf{Y} \equiv \begin{bmatrix}Y_1\\ \vdots\\ Y_N\end{bmatrix} \equiv \begin{bmatrix}Y(t_1)\\ \vdots\\ Y(t_N)\end{bmatrix} \equiv \begin{bmatrix}\ln(L(t_1))\\ \vdots\\ \ln(L(t_N))\end{bmatrix} \tag{7}$$

and a matrix of corresponding predictor variables

$$\begin{aligned}\vec{\mathbf{X}}_C &\equiv \begin{bmatrix}\mathbf{X}_1\\ \vdots\\ \mathbf{X}_N\end{bmatrix} \equiv \begin{bmatrix}\mathbf{X}(t_1)\\ \vdots\\ \mathbf{X}(t_N)\end{bmatrix} \\ &\equiv \begin{bmatrix}1 & \ln(Q(t_1)) & X_2(t_1) & \ldots & X_K(t_1)\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & \ln(Q(t_N)) & X_2(t_N) & \ldots & X_K(t_N)\end{bmatrix}\end{aligned} \tag{8}$$

ordinary least squares (OLS) regression estimates for the parameters of equation (4) are given by

$$\hat{\boldsymbol{\beta}} = \left(\vec{\mathbf{X}}_C'\vec{\mathbf{X}}_C\right)^{-1}\vec{\mathbf{X}}_C'\mathbf{Y} \tag{9}$$

$$s^2 = \left(\mathbf{Y} - \vec{\mathbf{X}}_C\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{Y} - \vec{\mathbf{X}}_C\hat{\boldsymbol{\beta}}\right)/(N-(K+1)) \tag{10}$$

The hats (e.g., "$\hat{\beta}$") identify estimators; unless indicated otherwise by subscript, variables with hats refer to maximum likelihood estimators.

### 2.4. Retransforming Rating Curve Estimates

[14] In general, the rating curve provides unbiased estimates of $\ln(L)$. Thus it is tempting to use the rating curve defined by equation (5), with parameters estimated by ordinary least squares, to estimate a continuous trace of the instantaneous loads by replacing $L(t)$ in equation (1) with

$$\hat{L}_{RC}(t) \equiv \exp\left(\mathbf{X}(t)\hat{\boldsymbol{\beta}}\right) \tag{11}$$

$$\equiv \exp\left(\hat{\beta}_0 + \hat{\beta}_1\ln(Q(t)) + \ldots\right) \tag{12}$$

$\hat{L}_{RC}$ is known as the rating curve load estimator. However, the statistical properties of $\hat{L}_{RC}$ are not ideal. $\hat{L}_{RC}$ exhibits so-called "retransformation bias" [*Miller*, 1984; *Thomas*, 1985; *Ferguson*, 1986]. Specifically, for a given discharge, $Q(t)$, the expected value of the true load is

$$E[L(t)|Q(t)] = \exp\left(\mathbf{X}(t)\beta + \sigma^2/2\right) \quad (13)$$

while the expected value of $\hat{L}_{RC}$ is

$$E\left[\hat{L}_{RC}(t)\right] = E\left[\exp\left(\mathbf{X}(t)\hat{\beta}\right)\right] \quad (14)$$

$$= \exp\left(\mathbf{X}(t)\beta + h_E(t)\sigma^2/2\right) \quad (15)$$

where

$$h_E(t) = \mathbf{X}(t)\left(\vec{\mathbf{X}}_C'\vec{\mathbf{X}}_C\right)^{-1}\mathbf{X}'(t) \quad (16)$$

$\hat{L}_{RC}$ is not just biased; it is also inconsistent. As sample size increases, $\hat{L}_{RC}$ does not converge to the expected value of the load. For large $N$, the ratio of the expected value of $\hat{L}_{RC}$ to $E[L]$ approaches $\exp(-\sigma^2/2)$. However, despite its bias, $\hat{L}_{RC}$ is important because it provides a basis for other estimators that are consistent [*Cohn et al.*, 1989]: (1)"quasi maximum likelihood estimator,"

$$\hat{L}_{QMLE} \equiv \hat{L}_{RC} \cdot \exp\left(s^2/2\right) \quad (17)$$

which is easy to employ and consistent though upwardly biased; (2) jackknife estimator,

$$\hat{L}_{JK} \equiv N \cdot \hat{L}_{QMLE} - \left(\frac{N-1}{N}\right)\sum_{i=1}^{N}\hat{L}_{Q_{[N-1],i}} \quad (18)$$

where $\hat{L}_{Q_{[N-1],i}}$ indicates $\hat{L}_{QMLE}$ with the $i$th observation (and corresponding predictor variables) deleted from the sample [*Efron*, 1982; *Stuart and Ord*, 1979]; (3) *Duan*'s [1983] "smearing estimator," a general approach for reducing retransformation bias applicable to many transformations including the logarithmic transformation employed here for which

$$\hat{L}_{Duan} \equiv \hat{L}_{RC} \cdot \sum_{i=1}^{N}\exp(\epsilon_i)/N \quad (19)$$

where

$$\epsilon_i \equiv Y_i - \mathbf{X}_i\hat{\beta}; \quad (20)$$

(4) Finney's minimum variance unbiased estimator (MVUE) [*Finney*, 1941; *Bradu and Mundlak*, 1970; *Stuart and Ord*, 1979],

$$\hat{L}_{MVUE} \equiv \hat{L}_{RC} \cdot g_m\left((1 - h_E(t))s^2/2\right) \quad (21)$$

where $m = (N - (K + 1))$ and $g_m(t) \equiv 1 + t + \frac{m}{m+2}\frac{t^2}{2!} + \frac{m^2}{(m+2)(m+4)}\frac{t^3}{3!} + \ldots$ (see section A5). $\hat{L}_{MVUE}$ is unbiased and, as a function of sufficient statistics, is also minimum variance for its expectation [*Finney*, 1941; *Aitchison and Brown*, 1981]. According to the standards of classical parametric estimation, $\hat{L}_{MVUE}$ completely solves the problem of load estimation with complete samples (see *Likěs* [1980] for additional discussion).

## 3. Estimation With Censored Data

[15] Employing data subject to type 1 censoring requires care. Both the censored and uncensored observations provide information about the loads, and threshold locations play an important role in determining the statistical properties of estimators regardless of whether corresponding sample observations are censored.

### 3.1. Notation

[16] Data subject to censoring can be characterized in various ways. In this paper, a pair of values, $\{Y_i, T_i\}$, is used to describe the $i$th observation, where $Y_i$ is the logarithm of the concentration (if $Y_i \geq T_i$) or a value less than $T_i$ (otherwise). In either case, $T_i$ corresponds to the logarithm of the threshold (in the orthophosphate example in section 2, $T_i \equiv \ln(0.02)$). It is also useful to define a "standardized censoring threshold" for each observation:

$$\xi_i = \frac{T_i - \mathbf{X}_i\beta}{\sigma} \quad (22)$$

Given the linear model, $\Phi(\xi_i)$ is the "censoring probability" corresponding to the ith observation.

### 3.2. Estimation

[17] It is not obvious how to use "less than" values in the context of OLS regression. In fact, *van Zwet* [1966] has shown that unbiased estimators are generally not available for type 1 censored data. However, given regularity conditions [*Rohatgi*, 1976, p. 361], maximum likelihood (ML) methods are often tractable (see *Dempster et al.* [1977] for discussion), asymptotically unbiased, and efficient for dealing with censored data [*Stuart and Ord*, 1979]. Furthermore, because the statistical properties of ML estimators can be derived analytically, ML provides a solid foundation for various adjustment methods [*Judge et al.*, 1985; *Stuart and Ord*, 1979, p. 74]. Finally, the ML and OLS estimators for $\beta$ are identical when dealing with complete data.

### 3.3. Likelihood Function

[18] The likelihood function for complete independent samples is proportional to the product of the likelihood (in this case, the probability density function, or *pdf*) corresponding to each observation:

$$\ell(\beta, \sigma) \propto \prod_{i=1}^{N} f_{\beta,\sigma}(Y_i) \quad (23)$$

where

$$f_{\beta,\sigma}(Y_i) \equiv \frac{e^{-(Y_i - \mathbf{X}_i\beta)^2/\left(2\sigma^2\right)}}{\sqrt{2\pi\sigma^2}} \quad (24)$$

is the probability density function (pdf) of $Y_i$ given $\mathbf{X}_i$ [*Stuart and Ord*, 1979].

[19] As with complete samples, the likelihood function for data subject to censoring is proportional to the product of the likelihood corresponding to each observation, but for the censored observations this is given by the cumulative distribution function (CDF) evaluated at the censoring threshold rather than by the pdf of the observed value:

$$\ell(\beta,\sigma) \propto \left\{ \prod_{Y_i|Y_i \geq T_i} f_{\beta,\sigma}(Y_i) \right\} \cdot \left\{ \prod_{Y_i|Y_i<T_i} F_{\beta,\sigma}(T_i) \right\} \tag{25}$$

where $F_{\beta,\sigma}(T_i)$ denotes the CDF of the normal distribution with parameters $(\mathbf{X_i}\beta, \sigma^2)$.

[20] The maximum likelihood parameter estimators, sometimes called "Tobit" estimators [*Tobin*, 1958] and usually denoted $\hat{\beta}$ and $\hat{\sigma}^2$, are those values of $\beta$ and $\sigma^2$ that maximize $\ell$. The maximum can be found by differentiating the logarithm of the likelihood function with respect to the parameters and solving the resulting $p = K + 2$ simultaneous nonlinear equations.

### 3.4. Moments of Maximum Likelihood Estimators

[21] *Shenton and Bowman* [1977] (hereinafter referred to as SB) provide general formulas for calculating the first-order bias ($\lim_{N\to\infty} N \cdot E[\hat{\theta}_i - \theta_i]$) and first-order covariance ($\lim_{N\to\infty} N \cdot E[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)]$) of ML estimators corresponding to distributions, like the normal and lognormal, that satisfy regularity conditions. For an arbitrary parameter $\theta_i$, with corresponding estimator $\hat{\theta}_i$, the first-order bias of $\theta_i$ can be expressed as a constant, $Bias_1[\hat{\theta}_i]$, defined by [*Shenton and Bowman*, 1977; *Hosking*, 1985; *Bowman and Shenton*, 1988]

$$\begin{aligned} Bias_1\left[\hat{\theta}_i\right] &\equiv \lim_{N\to\infty} N \cdot E\left[\hat{\theta}_i - \theta_i\right] \\ &= N \cdot \sum_{j=1}^{p} \sum_{k=1}^{p} \sum_{l=1}^{p} \psi^{ij}\psi^{kl}\left(\psi_{jk,l} + \frac{1}{2}\psi_{jkl}\right) \end{aligned} \tag{26}$$

and analogously the first-order covariances for $\hat{\theta}_i$ and $\hat{\theta}_j$ are defined by

$$\begin{aligned} Cov_1\left[\hat{\theta}_i, \hat{\theta}_j\right] &\equiv \lim_{N\to\infty} N \cdot E\left[\left(\hat{\theta}_i - \theta_i\right)\left(\hat{\theta}_j - \theta_j\right)\right] \\ &= N \cdot \psi^{ij} \end{aligned} \tag{27}$$

where

$$\begin{aligned} \psi_{jk} &= E\left[\frac{\partial^2 \ln(\ell)}{\partial\theta_j \partial\theta_k}\right] \\ \begin{bmatrix} \psi^{11} & \cdots & \psi^{1p} \\ \vdots & \vdots & \vdots \\ \psi^{p1} & \cdots & \psi^{pp} \end{bmatrix} &= \begin{bmatrix} \psi_{11} & \cdots & \psi_{1p} \\ \vdots & \vdots & \vdots \\ \psi_{p1} & \cdots & \psi_{pp} \end{bmatrix}^{-1} \\ \psi_{jkl} &= E\left[\frac{\partial^3 \ln(\ell)}{\partial\theta_j \partial\theta_k \partial\theta_l}\right] \\ \psi_{jk,l} &= E\left[\frac{\partial^2 \ln(\ell)}{\partial\theta_j \partial\theta_k} \cdot \frac{\partial \ln(\ell)}{\partial\theta_l}\right] \\ &= \frac{\partial}{\partial\theta_l}\psi_{jk} - \psi_{jkl} \end{aligned} \tag{28}$$

Identities for $\psi_{jk}$, $\psi_{jkl}$, and $\psi_{jk,l}$ appear in section A2. Equation (27) can be interpreted as Frechet-Cramér-Rao bounds, which are lower bounds on the variance of any unbiased estimator of the parameters [*Rohatgi*, 1976].

### 3.5. Required Threshold Information

[22] It is important to note that the formula for the first-order bias of ML estimators (equation (26)) depends on the censoring thresholds corresponding to each observation, even for those observations that exceed the censoring threshold. Obtaining this information can be a nuisance because many laboratories do not routinely report the thresholds corresponding to "detected" values.

[23] Given the thresholds corresponding to the observations, one then calculates $\hat{\xi}_i \equiv (T_i - \mathbf{X_i}\hat{\beta})/\hat{\sigma}$. Although using $\hat{\xi}_i$, which employs estimated parameters, instead of $\xi_i$ may seem troublesome from a theoretical viewpoint, in practice it has little impact on estimates.

[24] Standardized first-order biases and variances can be defined

$$\begin{aligned} B_{\hat{\beta}_i} &\equiv Bias_1\left[\hat{\beta}_i\right]/\sigma \\ B_{\hat{\sigma}} &\equiv Bias_1[\hat{\sigma}]/\sigma \\ B_{\hat{\sigma}^2} &\equiv Bias_1[\hat{\sigma}^2]/\sigma^2 \\ V_{\hat{\beta}_i\hat{\beta}_j} &\equiv Cov_1\left[\hat{\beta}_i, \hat{\beta}_j\right]/\sigma^2 \\ V_{\hat{\beta}_i\hat{\sigma}} &\equiv Cov_1\left[\hat{\beta}_i, \hat{\sigma}\right]/\sigma^2 \\ V_{\hat{\beta}_i\hat{\sigma}^2} &\equiv Cov_1\left[\hat{\beta}_i, \hat{\sigma}^2\right]/\sigma^3 \\ V_{\hat{\sigma}\hat{\sigma}} &\equiv Cov_1[\hat{\sigma}, \hat{\sigma}]/\sigma^2 \\ V_{\hat{\sigma}^2\hat{\sigma}^2} &\equiv Cov_1[\hat{\sigma}^2, \hat{\sigma}^2]/\sigma^4 \end{aligned}$$

that depend only on the censoring "pattern," the distribution of censoring thresholds corresponding to the observations, and not on the parameter values themselves. In the case of complete data, $B_{\hat{\beta}_i} = 0$ for all $i$, $B_{\hat{\sigma}^2} = -(K + 1)/N$, and $V_{\hat{\sigma}^2\hat{\sigma}^2} = 2/N$.

## 4. Load Estimation With Censored Data

### 4.1. Generalizing Complete-Data Estimators to Use Censored Data

[25] Censored-data load estimators can be readily obtained by inserting maximum likelihood censored-data parameter estimators $\{\hat{\beta}, \hat{\sigma}^2\}$ into the complete-data estimators defined in section 2.4. Although the censored-data versions of $\hat{L}_{RC}$ and $\hat{L}_{QMLE}$ (which is sometimes referred to as a Tobit load estimator [*Crawford*, 1996]) are easy to define, as is $\hat{L}_{JK}$, their statistical properties, specifically bias and variance, remain to be determined. The lack of residuals for censored observations precludes direct implementation of $\hat{L}_{Duan}$. For similar reasons, it is not immediately obvious how to adapt $\hat{L}_{MVUE}$ for censored data while preserving its desirable properties. The next section of this paper addresses that problem.

### 4.2. AMLE Load Estimator, $\hat{\mathbf{L}}_A$

[26] This section presents an adjusted maximum likelihood (AMLE) load estimator, $\hat{L}_A$, analogous to $\hat{L}_{MVUE}$. $\hat{L}_A$ is attractive both because of its theoretical underpinnings and, as will be demonstrated in Monte Carlo experiments,

because of its performance with heavy censoring and small samples [*Cohn et al.*, 1992b].

### 4.3. Reparameterization of the Regression Model

[27] AMLE load estimation begins with reparameterizing the linear model in terms of a new vector, denoted $\hat{\omega}$, defined in terms of $\hat{\beta}$ and $\hat{\sigma}$:

$$\hat{\boldsymbol{\omega}} = \begin{bmatrix} \hat{\omega}_0 \\ \hat{\omega}_1 \\ \dots \\ \hat{\omega}_K \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \dots \\ \hat{\beta}_K \end{bmatrix} - \left(\frac{1}{V_{\hat{\sigma}\hat{\sigma}}}\right) \begin{bmatrix} V_{\hat{\beta}_0\hat{\sigma}} \\ V_{\hat{\beta}_1\hat{\sigma}} \\ \vdots \\ V_{\hat{\beta}_K\hat{\sigma}} \end{bmatrix} \hat{\sigma} = \hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\hat{\sigma} \tag{29}$$

where $\boldsymbol{\gamma} \equiv (1/V_{\hat{\sigma}\hat{\sigma}}) \cdot \{V_{\hat{\beta}_0\hat{\sigma}}, \dots, V_{\hat{\beta}_K\hat{\sigma}}\}'$. The covariance of $\hat{\boldsymbol{\omega}}$ and $\hat{\sigma}$ is then

$$\begin{aligned} Cov[\hat{\boldsymbol{\omega}}, \hat{\sigma}] &= Cov\left[\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\hat{\sigma}, \hat{\sigma}\right] \\ &= Cov\left[\hat{\boldsymbol{\beta}}, \hat{\sigma}\right] - \boldsymbol{\gamma} Var[\hat{\sigma}] \\ &\approx Cov\left[\hat{\boldsymbol{\beta}}, \hat{\sigma}\right] - \frac{Cov\left[\hat{\boldsymbol{\beta}}, \hat{\sigma}\right]}{Var[\hat{\sigma}]} Var[\hat{\sigma}] \\ &= 0 \end{aligned} \tag{30}$$

### 4.4. Asymptotic Distribution of $\hat{\omega}$

[28] Under fairly general circumstances, the asymptotic distribution of $\hat{\beta}$ and $\hat{\sigma}$ is multivariate normal [*Robinson*, 1982], and therefore the distribution of $\hat{\omega}$ is approximately multivariate normal [*Kotz et al.*, 2000], $N[\mathbf{M}, \boldsymbol{\Sigma}]$, with parameters given by

$$\begin{aligned} \mathbf{M} &\equiv E[\hat{\boldsymbol{\omega}}] \\ &= E\left[\hat{\boldsymbol{\beta}}\right] - \boldsymbol{\gamma} E[\hat{\sigma}] \\ &\approx \boldsymbol{\beta} + \left(\vec{\mathbf{B}}/N\right)\sigma - \boldsymbol{\gamma}(1 + B_{\hat{\sigma}}/N)\sigma \\ &= \boldsymbol{\beta} + \vec{\mathbf{A}}\sigma \end{aligned} \tag{31}$$

where $\vec{\mathbf{B}} = \{B_{\hat{\beta}_0}, \dots B_{\hat{\beta}_K}\}'$ and $\vec{\mathbf{A}} \equiv \vec{\mathbf{B}}/N - \boldsymbol{\gamma}\,(1 + B_{\hat{\sigma}}/N)$; and

$$\begin{aligned} \boldsymbol{\Sigma} &\equiv Var[\hat{\boldsymbol{\omega}}] \\ &= Var\left[\hat{\boldsymbol{\beta}}\right] - 2 \cdot Cov\left[\hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}\hat{\sigma}\right] + Var[\boldsymbol{\gamma}\hat{\sigma}] \\ &= \sigma^2\left\{Var\left[\hat{\boldsymbol{\beta}}/\sigma\right] - Var[\boldsymbol{\gamma}\hat{\sigma}/\sigma]\right\} \\ &\approx \vec{\mathbf{C}} \cdot \sigma^2 \end{aligned} \tag{32}$$

where $\vec{\mathbf{C}} \equiv (\vec{\mathbf{V}} - \boldsymbol{\gamma}\boldsymbol{\gamma}' V_{\hat{\sigma}\hat{\sigma}})/N$, and

$$\vec{\mathbf{V}} = \begin{bmatrix} V_{\hat{\beta}_0\hat{\beta}_0} & \cdots & V_{\hat{\beta}_0\hat{\beta}_K} \\ \vdots & \dots & \vdots \\ V_{\hat{\beta}_K\hat{\beta}_0} & \cdots & V_{\hat{\beta}_K\hat{\beta}_K} \end{bmatrix} \tag{33}$$

Because $\hat{\omega}$ is (approximately) multivariate normal $N[\mathbf{M}, \boldsymbol{\Sigma}]$, for fixed

$$\mathbf{X} \equiv \{1, X_1, \dots, X_K\} \tag{34}$$

$\mathbf{X}\hat{\boldsymbol{\omega}}$ is univariate normal $N[\mathbf{XM}, \mathbf{X\Sigma X}'] \sim N[\mu_{\mathbf{X}\hat{\omega}}, \sigma^2_{\mathbf{X}\hat{\omega}}]$, where

$$\begin{aligned} \mu_{\mathbf{X}\hat{\omega}} &\approx \mathbf{X}\left(\boldsymbol{\beta} + \vec{\mathbf{A}}\sigma\right) \\ &= \mathbf{X}\boldsymbol{\beta} + \mathbf{X}\vec{\mathbf{A}}\sigma \\ \sigma^2_{\mathbf{X}\hat{\omega}} &\approx \mathbf{X}\vec{\mathbf{C}}\mathbf{X}'\sigma^2 \end{aligned} \tag{35}$$

### 4.5. Approximating the Small-Sample Distribution of $\hat{\sigma}^2$

[29] Although the exact properties of $\hat{\sigma}^2$ with censored data are unknown, *Cohn* [1988] found that the distribution of $\nu s_A^2/\sigma^2$, where $s_A^2 \equiv \frac{\hat{\sigma}^2}{\left(1+B_{\hat{\sigma}^2}/N\right)}$, is approximated well by a $\chi^2$ random variable with $\nu$ degrees of freedom where $\nu$ is given by

$$\nu \approx 2 \cdot N \cdot \left(\frac{\left(1 + B_{\hat{\sigma}^2}/N\right)^2}{V_{\hat{\sigma}^2\hat{\sigma}^2}}\right) \tag{36}$$

This approximation was found to be valid with censoring up to 80% and sample sizes as small as $N = 25$. In particular, for the case of complete samples, for which $B_{\hat{\sigma}^2} = -(K + 1)$, the approximation yields

$$\begin{aligned} s_A^2 &\equiv \left(\frac{N}{N - (K+1)}\right)\hat{\sigma}^2 \\ &= s^2 \end{aligned} \tag{37}$$

which is the familiar OLS estimator defined in equation (10), and the distribution of $(N - (K + 1))s^2/\sigma^2$ is known to be $\chi^2$ with $N - (K + 1)$ degrees of freedom [*Judge et al.*, 1985, p. 27, equation 2.1.19].

[30] Assuming that $\nu s_A^2/\sigma^2 \sim \chi^2_\nu$, a new function,

$$g_\nu\left(s_A^2, b_1, b_2\right) \equiv \sum_{k=0}^{\infty} c_k d_{k/2} \left(s_A^2\right)^{k/2} \tag{38}$$

can be defined where

$$d_k \equiv \left\{(2/\nu)^k \frac{\Gamma(\nu/2 + k)}{\Gamma(\nu/2)}\right\}^{-1} \tag{39}$$

$$c_k \equiv \sum_{n=0}^{\lfloor k/2} \left(\frac{b_2^n}{n!}\right)\left(\frac{b_1^{k-2n}}{(k-2n)!}\right) \tag{40}$$

and $\lfloor k/2$ indicates the greatest integer less than or equal to $k/2$. In practice, the infinite summation in equation (38) can usually be adequately approximated by the first 10 or fewer terms by stopping the summation once the magnitude of the next term becomes negligible.

[31] Given the assumptions, it can be shown (see section A5) that

$$E\left[g_\nu\left(s_A^2, b_1, b_2\right)\right] = \exp\left(b_1\sigma + b_2\sigma^2\right) \tag{41}$$

### 4.6. Estimator, $\hat{\mathbf{L}}_A$

[32] The adjusted maximum likelihood (AMLE) load estimator, $\hat{L}_A$, can now be defined:

$$\hat{L}_A \equiv \exp(\mathbf{X}\hat{\boldsymbol{\omega}}) \cdot g_\nu\left(s_A^2, -\mathbf{X}\vec{\mathbf{A}}, \left(1 - \mathbf{X}\vec{\mathbf{C}}\mathbf{X}'\right)/2\right) \tag{42}$$

Because $\hat{\sigma}^2$ (and therefore $s_A^2$) is asymptotically independent of $\hat{\omega}$,

$$E\left[\hat{L}_A\right] \approx E[\exp(\mathbf{X}\hat{\boldsymbol{\omega}})] \cdot E\left[g_\nu\left(s_A^2, -\mathbf{X}\vec{\mathbf{A}}, \left(1 - \mathbf{X}\vec{\mathbf{C}}\mathbf{X}'\right)/2\right)\right] \quad (43)$$

Assuming that $\mathbf{X}\hat{\boldsymbol{\omega}}$ is normally distributed with moments given in section 4.3.1 [*Loucks et al.*, 1981],

$$\begin{aligned} E[\exp(\mathbf{X}\hat{\boldsymbol{\omega}})] &= \exp\left(\mu_{\mathbf{X}\hat{\omega}} + \sigma^2_{\mathbf{X}\hat{\omega}}/2\right) \\ &= \exp\left(\mathbf{X}\beta + \mathbf{X}\vec{\mathbf{A}}\sigma + \mathbf{X}\vec{\mathbf{C}}\mathbf{X}'\sigma^2/2\right) \end{aligned} \quad (44)$$

and (see equation (A9)),

$$\begin{aligned} &E\left[g_\nu\left(s_A^2, -\mathbf{X}\vec{\mathbf{A}}, \left(1 - \mathbf{X}\vec{\mathbf{C}}\mathbf{X}'\right)/2\right)\right] \\ &\quad = \exp\left(-\mathbf{X}\vec{\mathbf{A}}\sigma + \left(1 - \mathbf{X}\vec{\mathbf{C}}\mathbf{X}'\right)\sigma^2/2\right) \end{aligned} \quad (45)$$

Substituting the results from equations (44) and (45) into equation (43) yields

$$E\left[\hat{L}_A\right] = \exp\left(\mathbf{X}\mathbf{b} + \sigma^2/2\right) \quad (46)$$

To the order of the approximations, $\hat{L}_A$ is an unbiased estimator for the mean load.

## 5. Computing Sums of Loads

[33] In computing annual or seasonal loads, one usually approximates the integral in equation (1) as a sum of loads corresponding to short intervals (e.g., days or hours):

$$\hat{L}_{year} = \sum_{day=1}^{365} \hat{L}_{year,day} = \sum_{day=1}^{365} \sum_{hour=1}^{24} \hat{L}_{year,day,hour} \quad (47)$$

or, more generally,

$$\hat{\mathbf{L}} = \sum_{i=1}^{N_e} \hat{L}_i \quad (48)$$

where the longer interval has been divided into $N_e$ subintervals. Because the expectation operator is linear, unbiasedness in the short-interval estimator ensures unbiasedness in the long-term estimator.

## 6. Uncertainty of AMLE Load Estimates

[34] Understanding the uncertainty in load estimates is important when communicating load information, in designing efficient sampling programs [*Gilroy et al.*, 1990a; *Vogel et al.*, 2003], and for other purposes.

[35] It is useful to define various quantities related to load, including **L**, the true and unknown load that was transported downstream; $\boldsymbol{\mu_L}$, the expected value of the true load, which can be understood as the conditional expectation of $L$ for a specific set of conditions; and $\hat{L}$, the estimated load.

[36] $\hat{L}$ is an estimate of $\boldsymbol{\mu_L}$, the mean of $\boldsymbol{L}$. Because $\boldsymbol{L}$ is unknowable without direct measurement, we employ $\hat{L}$ as a substitute for $\boldsymbol{L}$ with the assurance that, for an unbiased estimator, $E[\hat{L}] = E[\boldsymbol{L}] = \boldsymbol{\mu_L}$.

[37] However, it is still reasonable to ask how close the load estimate ($\hat{\mathbf{L}}$) is to the true load (**L**). This uncertainty can be described in terms of a "mean square error of prediction":

$$\begin{aligned} \mathrm{MSE}_P &\equiv E\left[\left(\mathbf{L} - \hat{\mathbf{L}}\right)^2\right] \\ &= Var\left[\mathbf{L} - \hat{\mathbf{L}}\right] + \left(E\left[\mathbf{L} - \hat{\mathbf{L}}\right]\right)^2 \end{aligned} \quad (49)$$

$$\begin{aligned} &= Var\left[\mathbf{L} - \hat{\mathbf{L}}\right] \\ &= Var[\mathbf{L} - \mu_L] + Var\left[\hat{\mathbf{L}} - \mu_L\right] \end{aligned} \quad (50)$$

The second term on the right-hand side of line 49 vanishes because the estimator is assumed to be unbiased, and on the final line $\mathrm{MSE}_P$ can be partitioned because $\hat{\mathbf{L}}$ is independent of $(\mathbf{L} - \mu_L)$.

[38] The two terms on the right-hand side of equation (50) refer, respectively, to two distinct sources of uncertainty: natural variability, or randomness that is essentially unpredictable, and parameter uncertainty, which is related to our limited understanding of the physical processes, specifically the uncertainty in parameter estimates.

[39] Parameter uncertainty can be reduced by increasing the number of samples, $N$, that are used to calibrate the regression model. Natural variability is a fundamental component of the natural process and cannot be reduced. Sections 6.1 and 6.3 address estimation of the two components of $\mathrm{MSE}_P$.

### 6.1. Natural Variability, $Var[\mathbf{L} - \mu_L]$

[40] The first term on the right-hand side of equation (50) can be expanded to yield

$$\begin{aligned} Var[\mathbf{L} - \boldsymbol{\mu_L}] &= Var\left[\sum_{i=1}^{N_e}\left(L_i - \mu_{L_i}\right)\right] \\ &= \sum_{j=1}^{N_e}\sum_{i=1}^{N_e} Cov\left[L_i, L_j\right] \\ &= \sum_{j=1}^{N_e}\sum_{i=1}^{N_e} \rho_{i,j} \cdot \sqrt{Var[L_i] \cdot Var\left[L_j\right]} \\ &= \sum_{j=1}^{N_e}\sum_{i=1}^{N_e} \rho_{i,j} \cdot \mu_i \cdot \mu_j \cdot \left(\exp\left(\sigma^2\right) - 1\right) \end{aligned} \quad (51)$$

where $\rho_{i,j} \equiv Cov[L_i, L_j]/\sqrt{Var[L_i] \cdot Var\left[L_j\right]}$ is the correlation between $\exp(\epsilon_i)$ and $\exp(\epsilon_j)$.

### 6.2. Estimating the Serial Correlation, $\rho_{i,j}$

[41] The assumption that $\epsilon(t)$ is independent of $\epsilon(t + \delta)$ is not generally valid for small values of $\delta$. The lag one day serial correlation in the errors typically exceeds 0.5, and sometimes 0.9 (G. Schwarz, personal communication, 2003), and varies with hydrological factors such as basin size, topography, and groundwater inflows, as well as by constituent. Moreover, we seldom have adequate data to estimate $\rho$ because estimating $\rho$ requires closely spaced data which are not well suited to estimating loads. Furthermore, because the standard deviation of $\hat{\rho}$ is approximately $\frac{1}{\sqrt{N}}$, it might take 100 or more observations to obtain acceptable precision; very few projects can afford to worry about it.

[42] For moderate sized watersheds (on the order of 1000 square miles) it is often assumed that the serial correlation is 1.0 within a day and 0.0 between days. This overestimates the within-day correlation and underestimates the between-day correlation. Fixing the value of ρ in this way, however, simplifies the computation, and, given the uncertainty in ρ, such an assumption may be reasonable. Also, it is important to recall that while the *uncertainty* in the load estimate may be sensitive to ρ, the *estimate* is insensitive to ρ because calibration data are collected infrequently. In addition, the relative magnitude of natural variability tends to diminish as the time interval over which loads are computed increases.

### 6.3. Parameter Uncertainty, $Var[\hat{\mathbf{L}} - \mu_L]$

[43] The second term on the right-hand side of equation (50) can be expanded to yield

$$Var\left[\hat{\mathbf{L}} - \mu_L\right] = Var\left[\sum_{i=1}^{N_e}\left(\hat{L}_{A,i} - \mu_{L_i}\right)\right] = \sum_{j=1}^{N_e}\sum_{i=1}^{N_e} Cov\left[\hat{L}_{A,i}, \hat{L}_{A,j}\right] \tag{52}$$

Thus the parameter uncertainty can be computed from the pairwise covariances. Let $\mathbf{X}_i$ and $\mathbf{X}_j$ represent two vectors of predictor variables corresponding to $\hat{L}_{A,i}$ and $\hat{L}_{A,j}$. The covariance of $\hat{L}_{A,i}$ and $\hat{L}_{A,j}$ is then given by

$$\begin{aligned} Cov\left[\hat{L}_{A,i}, \hat{L}_{A,j}\right] &= E\left[\left(\hat{L}_{A,i} - E\left[\hat{L}_{A,i}\right]\right)\cdot\left(\hat{L}_{A,j} - E\left[\hat{L}_{A,j}\right]\right)\right] \\ &= E\left[\hat{L}_{A,i}\cdot\hat{L}_{A,j}\right] - E\left[\hat{L}_{A,i}\right]\cdot E\left[\hat{L}_{A,j}\right] \\ &= E\left[\hat{L}_{A,i}\cdot\hat{L}_{A,j}\right] - \exp\left((\mathbf{X}_i + \mathbf{X}_j)\beta + \sigma^2\right) \end{aligned} \tag{53}$$

Because $\hat{\omega}$ is asymptotically orthogonal to $s_A^2$, the first term on the right-hand side of equation (53) can be expanded into

$$\begin{aligned} E\left[\hat{L}_{A,i}\cdot\hat{L}_{A,j}\right] &= E\Big[\exp(\mathbf{X}_i\hat{\omega})g_\nu\left(s_A^2, -\mathbf{X}_i\vec{\mathbf{A}}, \left(1 - \mathbf{X}_i\vec{\mathbf{C}}\mathbf{X}'\right)/2\right) \\ &\quad \cdot \exp(\mathbf{X}_j\hat{\omega})g_\nu\left(s_A^2, -\mathbf{X}_j\vec{\mathbf{A}}, \left(1 - \mathbf{X}_j\vec{\mathbf{C}}\mathbf{X}_j'\right)/2\right)\Big] \\ &\approx E[F_1]\cdot E[F_2] \end{aligned} \tag{54}$$

where

$$F_1 = \exp\left((\mathbf{X}_i + \mathbf{X}_j)\hat{\omega}\right) \tag{55}$$

and

$$\begin{aligned} F_2 &= g_\nu\left(s_A^2, -\mathbf{X}_i\vec{\mathbf{A}}, \left(1 - \mathbf{X}_i\vec{\mathbf{C}}\mathbf{X}_i'\right)/2\right) \\ &\quad \cdot g_\nu\left(s_A^2, -\mathbf{X}_j\vec{\mathbf{A}}, \left(1 - \mathbf{X}_j\vec{\mathbf{C}}\mathbf{X}_j'\right)/2\right) \end{aligned} \tag{56}$$

[44] Assuming $\mathbf{X}_i\hat{\omega}$ is normally distributed with moments given in section 4.3.1, the expected value of $F_1$ is

$$\begin{aligned} E[F_1] &= E\left[\exp\left((\mathbf{X}_i + \mathbf{X}_j)\cdot\hat{\omega}\right)\right] \\ &= \exp\Big((\mathbf{X}_i + \mathbf{X}_j)\beta + (\mathbf{X}_i + \mathbf{X}_j)\vec{\mathbf{A}}\sigma \\ &\quad + (\mathbf{X}_i + \mathbf{X}_j)\vec{\mathbf{C}}(\mathbf{X}_i + \mathbf{X}_j)'\sigma^2/2\Big) \end{aligned} \tag{57}$$

[45] $F_2$ involves the product of two $g_\nu(\cdot)$ functions. The expected value of $F_2$, denoted $G_\nu()$, can be evaluated by considering the series expansions:

$$\begin{aligned} &G_\nu\left(\sigma^2, b_{11}, b_{12}, b_{21}, b_{22}\right) \\ &\quad\equiv E[F_2] \\ &\quad\equiv E\left[g_\nu\left(s_A^2, b_{11}, b_{12}\right)\cdot g_\nu\left(s_A^2, b_{21}, b_{22}\right)\right] \\ &\quad= E\left[\left\{\sum_{k=0}^{\infty} c_{1,k}d_k s_A^k\right\}\cdot\left\{\sum_{k=0}^{\infty} c_{2,k}d_k s_A^k\right\}\right] \\ &\quad= \sum_{j=0}^{\infty} E\left[s_A^j\right]\sum_{k=0}^{j}\left(c_{1,k}d_k\right)\cdot\left(c_{2,j-k}d_{j-k}\right) \\ &\quad= \sum_{j=0}^{\infty}\left\{\sigma^j/d_j\right\}\sum_{k=0}^{j}\left(c_{1,k}d_k\right)\cdot\left(c_{2,j-k}d_{j-k}\right) \end{aligned} \tag{58}$$

Putting the pieces together yields

$$\begin{aligned} Cov\left[\hat{L}_{A,i}, \hat{L}_{A,j}\right] &= \exp\Big((\mathbf{X}_i + \mathbf{X}_j)\beta + (\mathbf{X}_i + \mathbf{X}_j)\vec{\mathbf{A}}\sigma \\ &\quad + (\mathbf{X}_i + \mathbf{X}_j)\vec{\mathbf{C}}(\mathbf{X}_i + \mathbf{X}_j)'\sigma^2/2) \\ &\quad \cdot G_\nu\left(\sigma^2, -\mathbf{X}_i\vec{\mathbf{A}}, \left(1 - \mathbf{X}_i\vec{\mathbf{C}}\mathbf{X}_i'\right)/2, -\mathbf{X}_j\vec{\mathbf{A}},\right. \\ &\quad \left. \cdot\left(1 - \mathbf{X}_j\vec{\mathbf{C}}\mathbf{X}_j'\right)/2\right) - \exp\left((X_i + X_j)\beta + \sigma^2\right) \end{aligned} \tag{59}$$

The result in equation (59) can be inserted into equation (52) to compute the parameter uncertainty. $MSE_P$, as defined in equation (50), can then be computed as the sum of the natural variability and parameter uncertainty.

[46] Note that the parameter uncertainty, $Var[\hat{\mathbf{L}} - \mu_L]$, will vary according to the estimator employed; the results developed above are specific to $\hat{L}_A$ (*Gilroy et al.* [1990a] provide equations for parameter uncertainty corresponding to $\hat{L}_{RC}$, $\hat{L}_{QMLE}$, $\hat{L}_{Duan}$, and $\hat{L}_{MVUE}$, for the complete data case). Natural variability, $Var[\mathbf{L} - \mu_L]$, does not depend on which load estimator is used.

### 6.4. Approximate Confidence Intervals

[47] Given $\hat{L}_A$ and $MSE_p$, approximate 95% confidence intervals for the true load can be obtained in at least two ways.

[48] 1. Normal assumption

$$\begin{aligned} \sigma_N &= \sqrt{MSE_p} \\ \mu_N &= \hat{L}_A \\ CI_N &= \{\mu_N \pm 1.96\cdot\sigma_N\} \end{aligned} \tag{60}$$

[49] 2. Lognormal assumption

$$\begin{aligned} \sigma_{LN} &= \sqrt{\ln\left(1 + MSE_p/\hat{L}_A^2\right)} \\ \mu_{LN} &= \ln\hat{L}_A - \sigma_{LN}^2/2 \\ CI_{LN} &= \{\exp(\mu_{LN} \pm 1.96\cdot\sigma_{LN})\} \end{aligned} \tag{61}$$

[50] It appears that confidence intervals based on the lognormal assumption may be more accurate than those based on the normal assumption in the sense that the

**Table 1.** Bias of Load Estimators for $T = 0$

| | | Estimator | | | |
|---|---|---|---|---|---|
| $N$ | $\sigma^2$ | $\hat{L}_{RC}$ | $\hat{L}_{QMLE}$ | $\hat{L}_{JK}$ | $\hat{L}_A$ |
| 20 | 0.04 | −1.4 | 0.4 | −0.4 | 0.0 |
| 40 | 0.04 | −1.7 | 0.2 | −0.1 | 0.0 |
| 80 | 0.04 | −1.8 | 0.1 | 0.0 | 0.0 |
| 20 | 0.25 | −8.8 | 2.2 | −1.4 | 0.0 |
| 40 | 0.25 | −10.3 | 1.1 | −0.2 | 0.0 |
| 80 | 0.25 | −11.1 | 0.5 | 0.0 | 0.0 |
| 20 | 0.64 | −21.6 | 5.8 | −3.2 | 0.0 |
| 40 | 0.64 | −24.7 | 2.8 | −0.6 | 0.0 |
| 80 | 0.64 | −26.1 | 1.4 | −0.1 | 0.0 |
| 20 | 1.44 | −42.8 | 16.4 | −12.2 | −0.2 |
| 40 | 1.44 | −47.4 | 7.5 | −1.7 | 0.0 |
| 80 | 1.44 | −49.5 | 3.7 | −0.3 | 0.1 |

probability of failure to cover in each tail is closer to 2.5%. However, this remains to be verified.

## 7. Monte Carlo Experiments

[51] Monte Carlo experiments were conducted to explore and compare the small-sample performance of $\hat{L}_{RC}$, $\hat{L}_{QMLE}$, $\hat{L}_{JK}$, and $\hat{L}_A$, and to verify that $\hat{L}_A$ performs as expected based on the analytical results in section 6.

### 7.1. Simulated Populations

[52] One hundred thousand replicate samples were generated, with samples of size 20, 40, and 80, corresponding to a regression model containing a constant and one nonconstant predictor variable intended to represent ln($Q$). The predictor variable was assigned values of $\ln(Q_i) = \Phi^{-1}((i - 0.63)/(N + 1 - 2 \cdot 0.63)$, where $\Phi$ refers to the standard normal CDF (0.63 ensures that the variance of the log flows is close to 1.0). The coefficient corresponding to ln($Q$), $\beta_1$, was set equal to 0.5 for the concentration model, a typical value for phosphorus-related compounds, and $\beta_0$ was set equal to 0.0, which entails no loss of generality.

[53] Because the behavior of the load estimators has been found to be sensitive to the skewness of the real-space error distribution (which is a function of $\sigma^2$), experiments were conducted with $\sigma^2 = \{0.04, 0.25, 0.64, 1.44\}$, which covers the range of practical interest. This corresponds to coefficients of variation of {0.20, 0.53, 0.95, 1.79} and coefficients of skewness of {0.6, 1.8, 3.7, 11.2}, respectively.

[54] Experiments were run with various censoring thresholds. The thresholds were applied to the logarithms of the concentration data at fixed values of $T = \{-\infty, -2\sigma, -\sigma, 0, \sigma\}$. In addition, three observations in each sample, specifically those concentrations corresponding to the highest discharges, were never censored (equivalent to thresholds of $T = -\infty$) to ensure that in all cases there would be sufficient data to obtain a fit. The overall percentage of observations censored ranged from 0%, for the case $T = -\infty$, to more than 80%, for $T = \sigma$.

[55] The estimators' performance was judged in terms of ability to estimate the expected sum of loads corresponding to $N_e = 20$ fixed discharges that were distributed in the same manner as the discharges used for calibration for the case $N = 20$. The mean and variance of the sum of the $N_e = 20$ estimated loads were used to compute the percentage bias and standard deviation of each estimator, and these two criteria were used to judge the overall performance.

[56] Table 1 reports the percent bias of $\hat{L}_{RC}$, $\hat{L}_{QMLE}$, $\hat{L}_{JK}$, and $\hat{L}_A$ for the case $T = 0$. For this case, approximately half of all observations were reported as censored values. $\hat{L}_{RC}$ is, as expected, downward biased in all cases. $\hat{L}_{QMLE}$ performs well with large samples, but is upward biased with smaller sample sizes. The jackknife estimator, $\hat{L}_{JK}$, exhibits high bias for small sample sizes, but the bias declines rapidly as sample size increases. $\hat{L}_A$ exhibited no significant bias in any of the 12 experiments.

#### 7.1.1. Standard Deviations of Load Estimators

[57] Table 2 reports the percent standard deviations of $\hat{L}_{RC}$, $\hat{L}_{QMLE}$, $\hat{L}_{JK}$, and $\hat{L}_A$ for the same cases considered in Table 1. Except in the extreme case where $\sigma^2 = 1.44$, the differences in the standard deviations can be attributed almost entirely to the differences in bias (the standard deviation is generally proportional to the expected value for lognormal variates, and for this reason a downward bias results in a similar reduction in standard deviation). The negatively biased estimators, $\hat{L}_{RC}$ and $\hat{L}_{JK}$, exhibit the lowest standard deviations, in some cases below the first-order results which correspond to FCR bounds. $\hat{L}_{QMLE}$, which is positively biased, exhibited considerably larger standard deviations.

#### 7.1.2. Effect of Censoring Threshold on AMLE Performance

[58] Table 3 reports the percent bias of $\hat{L}_A$ for 60 cases encompassing combinations of $N = \{20, 40, 80\}$, $\sigma^2 = \{0.04, 0.25, 0.64, 1.44\}$, and $T = \{-\infty, -2\sigma, -\sigma, 0, \sigma\}$. $\hat{L}_A$ exhibits significant ($\alpha = 5\%$ test) bias in two of the cases, with a maximum magnitude amounting to 0.5% of the estimate. Table 4 reports the percent standard deviation of $\hat{L}_A$ ("Obs") for the 60 cases presented in Table 3, as well as the average of the estimated analytical standard deviations derived in section 6 ("Est"). The observed standard deviation of $\hat{L}_A$ is generally close to the analytical estimates based on equation (52). Because the first-order results also define the FCR bounds, this indicates that $\hat{L}_A$ performs about as well as is possible for any unbiased estimator.

### 7.2. Discussion of Monte Carlo Results

[59] The load estimators all performed similarly in terms of variance. However, they differed substantially with respect to bias. As expected, $\hat{L}_{QMLE}$ is upwardly biased, and $\hat{L}_{RC}$ and $\hat{L}_{JK}$ are generally negatively biased. The AMLE, $\hat{L}_A$, exhibits vanishingly little bias.

**Table 2.** Standard Deviation of Load Estimators for $T = 0$: Parameter Uncertainty

| | | Estimator | | | |
|---|---|---|---|---|---|
| $N$ | $\sigma^2$ | $\hat{L}_{RC}$ | $\hat{L}_{QMLE}$ | $\hat{L}_{JK}$ | $\hat{L}_A$ |
| 20 | 0.04 | 8.1 | 8.3 | 8.6 | 8.2 |
| 40 | 0.04 | 5.7 | 5.9 | 5.9 | 5.8 |
| 80 | 0.04 | 4.0 | 4.1 | 4.1 | 4.1 |
| 20 | 0.25 | 18.5 | 21.7 | 21.0 | 20.8 |
| 40 | 0.25 | 12.7 | 15.0 | 14.6 | 14.6 |
| 80 | 0.25 | 8.9 | 10.4 | 10.3 | 10.3 |
| 20 | 0.64 | 26.1 | 39.8 | 36.6 | 35.0 |
| 40 | 0.64 | 17.2 | 25.9 | 24.2 | 24.4 |
| 80 | 0.64 | 11.8 | 17.6 | 17.1 | 17.1 |
| 20 | 1.44 | 30.3 | 89.0 | 167.6 | 59.6 |
| 40 | 1.44 | 18.5 | 47.7 | 39.5 | 40.7 |
| 80 | 1.44 | 12.3 | 30.3 | 28.0 | 28.2 |

**Table 3.** Percent Bias of AMLE Load Estimator

| | | Censoring Threshold | | | | |
|---|---|---|---|---|---|---|
| $N$ | $\sigma^2$ | $T = -\infty$ | $T = -2\sigma$ | $T = -\sigma$ | $T = 0$ | $T = \sigma$ |
| 20 | 0.04 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 40 | 0.04 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 80 | 0.04 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 20 | 0.25 | 0.0 | 0.0 | 0.1 | 0.0 | −0.4 |
| 40 | 0.25 | 0.0 | 0.0 | 0.0 | 0.0 | −0.1 |
| 80 | 0.25 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 20 | 0.64 | 0.0 | 0.0 | 0.0 | 0.0 | −0.5 |
| 40 | 0.64 | 0.0 | 0.0 | 0.0 | 0.0 | −0.2 |
| 80 | 0.64 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 20 | 1.44 | −0.1 | −0.1 | −0.1 | −0.2 | −0.3 |
| 40 | 1.44 | 0.0 | −0.1 | −0.1 | 0.0 | −0.2 |
| 80 | 1.44 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 |

[60] While the 60 Monte Carlo experiments presented here do not fully cover all circumstances one might encounter in practice, they do confirm the validity of the analytical results (section 6) over a range of conditions and provide confidence that the estimators' performance with more complicated models (such as the seven-parameter model of *Cohn et al.* [1992a]) can be characterized by analytical methods, even for fairly heavy censoring and small to moderate sample sizes.

## 8. Example: Orthophosphate Loads

[61] The following section shows how the previously described techniques can be applied to estimate the average daily load of a nutrient at a site, in this case orthophosphate (as $P$) carried by the Susquehanna River at Conowingo, Maryland (USGS gauge number 01578310) during water year 2002 (10/1/2001 to 9/30/2002). The data used to calibrate and run the model can be retrieved from NWISWeb Data for the Nation (http://waterdata.usgs.gov/nwis/). The calibration data includes orthophosphate concentration measurements (USGS parameter code 00671) and the record of mean daily discharge at the site collected during water years 1994–2003 (it has been found that limiting the calibration data to a 10-year moving window, where final loads are computed for year 9 within the window, yields reasonably timely and precise estimates while protecting against nonstationarity in the process). The complete data set includes 257 observations, 43 of which are reported as less than values.

### 8.1. Fitting the Seven-Parameter Model

[62] A seven-parameter model incorporating predictors for flow dependence, time trends, and seasonality [*Cohn et al.*, 1992a] was fit by maximum likelihood to the $N$ = 257 observations, yielding a prediction equation for the logarithm of load:

$$\begin{aligned}\ln(L) = 6.266 + 1.178 \cdot \ln(Q/\bar{Q}) + 0.083 \cdot (\ln(Q/\bar{Q}))^2 \\ - 0.023 \cdot (T - \bar{T}) + 0.032 \cdot (T - \bar{T})^2 \\ - 0.093 \cdot \sin(2 * \pi * T) + 0.445 \cdot \cos(2 * \pi * T) + error\end{aligned} \quad (62)$$

The estimated variance of the residual error was $\hat{\sigma}^2 = 0.965^2 = 0.930$, with a lag one (corresponding to an average lag time of about 2 weeks) serial correlation of 0.03. $\bar{Q}$ = 42024.16 and $\bar{T}$ = 1998.533 are centering variables (designed to ensure orthogonality between the linear and quadratic predictors corresponding to ln($Q$) and $T$).

### 8.2. Estimating Loads

[63] Using daily discharge data for water year 2002, the daily average orthophosphate load was estimated using the four load estimators, $\hat{L}_{RC}$, $\hat{L}_{QMLE}$, $\hat{L}_{JK}$, and $\hat{L}_A$. In order to explore the effect of sample size, the seven-parameter model was fit and loads were calculated using subsets of the $N$ = 257 calibration observations. Specifically, results were computed with all of the data ($N$ = 257), with every second observation ($N$ = 128), with every fourth observation ($N$ = 64), and with every eighth observation ($N$ = 32).

[64] The results of the subsampling experiment are reported in Table 5. With the exception of $\hat{L}_{RC}$, which exhibited substantial downward bias, the differences among the estimates is relatively small, on the order of 1% or less. This result is entirely consistent with what was observed in the Monte Carlo experiments (see Table 1).

### 8.3. Estimating the Uncertainty

[65] Assuming that the lag one day serial correlation ($\rho$) is zero between days and $\rho$ = 1 within days, the uncertainty in $\hat{L}_A$ can be evaluated using equations (51) and (59). Table 6 presents sources of uncertainty and lognormal confidence intervals (equation (61)) around the estimated load, $\hat{L}_A$, for calibration data sets ranging from $N$ = 32 to $N$ = 257 observations. As expected, the estimated natural variability is relatively constant, while the parameter uncertainty is approximately inversely proportional to $N$. However, for $N$ < 64, the parameter uncertainty grows dramatically, possibly suggesting that in practice sample sizes should be at least that large. It is also interesting to note that in all cases the estimated parameter uncertainty greatly exceeds

**Table 4.** Percent Standard Deviation of AMLE Load Estimator[a]

| | | Censoring Threshold | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $T = -\infty$ | | $T = -2\sigma$ | | $T = -\sigma$ | | $T = 0$ | | $T = \sigma$ | |
| $N$ | $\sigma^2$ | Obs | Est | Obs | Est | Obs | Est | Obs | Est | Obs | Est |
| 20 | 0.04 | 7.4 | 7.0 | 7.9 | 7.5 | 8.1 | 7.7 | 8.2 | 7.8 | 8.4 | 7.8 |
| 40 | 0.04 | 5.2 | 5.1 | 5.6 | 5.5 | 5.7 | 5.6 | 5.8 | 5.7 | 5.9 | 5.7 |
| 80 | 0.04 | 3.7 | 3.7 | 4.0 | 3.9 | 4.0 | 4.0 | 4.1 | 4.1 | 4.2 | 4.1 |
| 20 | 0.25 | 19 | 19 | 19 | 19 | 20 | 20 | 21 | 21 | 22 | 21 |
| 40 | 0.25 | 13 | 13 | 14 | 14 | 14 | 14 | 15 | 15 | 15 | 15 |
| 80 | 0.25 | 9.4 | 9.4 | 9.6 | 9.6 | 9.9 | 9.9 | 10 | 10 | 11 | 10 |
| 20 | 0.64 | 32 | 32 | 32 | 33 | 33 | 35 | 35 | 40 | 45 | 40 |
| 40 | 0.64 | 22 | 22 | 22 | 23 | 23 | 24 | 24 | 26 | 27 | 26 |
| 80 | 0.64 | 16 | 16 | 16 | 16 | 16 | 17 | 17 | 18 | 18 | 18 |
| 20 | 1.44 | 52 | 59 | 53 | 61 | 55 | 67 | 60 | 84 | 167 | 84 |
| 40 | 1.44 | 36 | 38 | 36 | 39 | 38 | 42 | 41 | 48 | 46 | 48 |
| 80 | 1.44 | 25 | 26 | 25 | 26 | 26 | 28 | 28 | 31 | 32 | 31 |

[a]Obs, observed; Est, estimated.

**Table 5.** Estimates of Average Daily Orthophosphate Loads, Susquehanna River at Conowingo, Maryland, WY2002

| Calibration Data | | Estimator, kg/d | | | |
|---|---|---|---|---|---|
| Total, $N$ | "Less Thans" | $\hat{L}_{RC}$ | $\hat{L}_{QMLE}$ | $\hat{L}_{JK}$ | $\hat{L}_A$ |
| 257 | 43 | 456.5 | 717.5 | 720.2 | 720.2 |
| 128 | 21 | 492.0 | 759.6 | 764.0 | 763.2 |
| 64 | 9 | 485.1 | 747.7 | 759.0 | 755.6 |
| 32 | 5 | 631.3 | 951.4 | 965.5 | 963.3 |

**Table 6.** Uncertainty in AMLE Estimates of Average Daily Orthophosphate Loads, Susquehanna River at Conowingo, Maryland, WY2002

| | Estimate, kg/d | | Uncertainty, $kg^2/d^2$ | | |
|---|---|---|---|---|---|
| $N$ | $\hat{L}_A$ | $CI_{LN}$ (95%) | $MSE_p$ | Natural | Parameter |
| 257 | 720.2 | (528, 960) | 12,187 | 4,850 | 7,337 |
| 127 | 763.2 | (517, 1087) | 21,309 | 5,333 | 15,976 |
| 64 | 755.6 | (452, 1189) | 35,873 | 5,649 | 30,224 |
| 32 | 963.3 | (473, 1755) | 110,060 | 11,056 | 99,004 |

the natural variability, indicating that the total uncertainty in the load estimate, even with $N$ = 257 observations, can be substantially reduced by increasing the number of observations used for calibration.

## 9. Future Research

[66] Work remains to be done in several areas. The first is natural variability. The short-term correlation structure of the error term needs to be explored in detail. Real-time sensors [*Christensen and Ziegler*, 2000], which can provide the required data at modest cost, should help to improve our understanding of the time series structure of concentration data.

[67] The second area is CI formulas. The accuracy of the confidence interval formulas needs to be verified, and the formulas may need to be improved.

[68] The third area is robustness. The robustness of the AMLE estimator with data that deviate from the hypothesized linear model remains to be explored. It is believed, however, that censoring, which in this case has the effect of eliminating low outliers, will simultaneously remove difficulties with model misspecification. Thus the validity of the linear model (equation (2)) with complete samples [*Cohn et al.*, 1992a; *Crawford*, 1996] may be relied on to justify its use with censored data.

## 10. Conclusions

[69] The AMLE method presented here is a natural extension of the well-accepted MVUE for complete data, and it shares essentially all of the MVUE's desirable properties. Monte Carlo experiments with moderate-sized samples show that the AMLE exhibits negligible bias and performs better than any of the alternative estimators that have been tested. In fact, the AMLE comes very close to achieving Frechet-Cramér-Rao bounds on its variance; by the standards of classical estimation, the AMLE is essentially optimal.

[70] The statistical framework presented here, which involves partitioning overall uncertainty of load estimates into two components, explicitly recognizes and quantifies the origins of uncertainty in load estimates. For both censored and complete data, the framework should facilitate cost-effective monitoring of nutrient transport and improve understanding of the uncertainty in load estimates.

## Appendix A

### A1. Availability of Software

[71] FORTRAN software to implement the AMLE methods is available from the author. The AMLE estimation method is also available as part of the LOADEST package [*Runkel et al.*, 2004], which can be downloaded from the Web at http://water.usgs.gov/pubs/tm/2004/tm4A5/.

### A2. Properties of the Likelihood Function

[72] The terms in equations (26)–(28), which are used to compute the bias and variance of the parameter estimators, are expected values of the partial derivatives of the logarithm of the likelihood ($\ln(\ell)$) function (equation (25)) with respect to the model parameters $\boldsymbol{\beta}$ and $\sigma$.

[73] We will first consider the case $\vec{\mathbf{X}}_C = \{1, 1, \ldots, 1\}'$, and $\boldsymbol{\beta} = \{\beta_0\} = \mu$ (the general case, discussed in section A3, is a simple extension). Following the approach of *Cohn* [1988], partial derivatives of the log likelihood function are taken with respect to $\mu$ and $\sigma$, and their expectations are computed. The following identities are useful in deriving the following terms. (1) The overall log likelihood is the sum of the log likelihoods corresponding to each observation. Similarly,

$$\psi_* = \sum_{i=1}^{N} \psi_*(i) \tag{A1}$$

where $\psi_*(i)$ refers to the partial derivative terms corresponding to observation $i$. (2) The conditional random variable $Y_i|Y_i \geq T_i$ is (trivially) independent of whether $Y_i \geq T_i$. (3) $P[Y_i \geq T_i] = 1 - \Phi[\xi_i]$. (4) The moments of a censored standard normal deviate, $Z$, can be obtained by recursion:

$$\begin{aligned} E\left[Z^0|Z \geq \xi\right] &= 1 \\ E\left[Z^1|Z \geq \xi\right] &= \frac{\phi}{1-\Phi} \\ E\left[Z^k|Z \geq \xi\right] &= (k-1)\cdot E\left[Z^{k-2}|Z \geq \xi\right] + \left(\frac{\phi}{1-\Phi}\right)\cdot \xi^{k-1} \end{aligned} \tag{A2}$$

[74] Expressions appear below for the value of $\psi_*(i)$ corresponding to a single observation for the case, described above, of a single predictor variable, $\mu$, and the error standard deviation, $\sigma$. To simplify the equations, subscripts have been omitted where there is no ambiguity.

$$\begin{aligned} \psi_{11} = \psi_{\mu\mu} &= \frac{\left(-\phi^2 - \Phi - \xi\phi\Phi + \Phi^2\right)}{\sigma^2\Phi} \\ \psi_{12} = \psi_{\mu\sigma} &= -\frac{\phi\left(\xi\phi + \Phi + \xi^2\Phi\right)}{\sigma^2\Phi} \\ \psi_{21} = \psi_{\sigma\mu} &= \psi_{\mu\sigma} \\ \psi_{22} = \psi_{\sigma\sigma} &= \frac{-\left(\xi^2\phi^2\right) - 2\Phi - \xi\phi\Phi - \xi^3\phi\Phi + 2\Phi^2}{\sigma^2\Phi} \end{aligned} \tag{A3}$$

$$\begin{aligned} \psi_{\mu\mu\mu} &= -\phi\left(2\phi^2 + 3\xi\phi\Phi - \Phi^2 + \xi^2\Phi^2\right)/\left(\sigma^3\Phi^2\right) \\ \psi_{\mu\mu\sigma} &= -\left(2\xi\phi^3 - 2\phi^2\Phi + 3\xi^2\phi^2\Phi - 2\Phi^2 - 3\xi\phi\Phi^2 \right. \\ &\quad \left. + \xi^3\phi\Phi^2 + 2\Phi^3\right)/\left(\sigma^3\Phi^2\right) \\ \psi_{\mu\sigma\mu} &= \psi_{\sigma\mu\mu} = \psi_{\mu\mu\sigma} \\ \psi_{\mu\sigma\sigma} &= -\phi\left(2\xi^2\phi^2 - 4\xi\phi\Phi + 3\xi^3\phi\Phi - 4\Phi^2 - 5\xi^2\Phi^2 + \xi^4\Phi^2\right) \\ &\quad /\left(\sigma^3\Phi^2\right) \\ \psi_{\sigma\mu\sigma} &= \psi_{\sigma\sigma\mu} = \psi_{\mu\sigma\sigma} \\ \psi_{\sigma\sigma\sigma} &= -\left(2\xi^3\phi^3 - 6\xi^2\phi^2\Phi + 3\xi^4\phi^2\Phi - 10\Phi^2 - 6\xi\phi\Phi^2 - 7\xi^3\phi\Phi^2 \right. \\ &\quad \left. + \xi^5\phi\Phi^2 + 10\Phi^3\right)/\left(\sigma^3\Phi^2\right) \end{aligned} \tag{A4}$$

$$
\begin{aligned}
\psi_{\mu\mu,\mu} &= -\frac{\phi\left(-\phi^2 - \xi\phi\Phi + \Phi^2\right)}{\sigma^3\Phi^2} \\
\psi_{\mu\mu,\sigma} &= -\frac{\xi\phi\left(-\phi^2 - \xi\phi\Phi + \Phi^2\right)}{\sigma^3\Phi^2} \\
\psi_{\mu\sigma,\mu} &= \frac{\left(\xi\phi^3 - \phi^2\Phi + \xi^2\phi^2\Phi - 2\Phi^2 - 2\xi\phi\Phi^2 + 2\Phi^3\right)}{\sigma^3\Phi^2} \\
\psi_{\sigma\mu,\mu} &= \psi_{\mu\sigma,\mu} \\
\psi_{\mu\sigma,\sigma} &= -\frac{\phi(-(\xi\phi) + 2\Phi)\left(\xi\phi + \Phi + \xi^2\Phi\right)}{\sigma^3\Phi^2} \\
\psi_{\sigma\mu,\sigma} &= \psi_{\mu\sigma,\sigma} \\
\psi_{\sigma\sigma,\mu} &= -\frac{\phi\left(-\left(\xi^2\phi^2\right) + 2\xi\phi\Phi - \xi^3\phi\Phi + 5\Phi^2 + 3\xi^2\Phi^2\right)}{\sigma^3\Phi^2} \\
\psi_{\sigma\sigma,\sigma} &= \left(\xi^3\phi^3 - 2\xi^2\phi^2\Phi + \xi^4\phi^2\Phi - 6\Phi^2 - 5\xi\phi\Phi^2 \right. \\
&\quad \left. - 3\xi^3\phi\Phi^2 + 6\Phi^3\right)/\left(\sigma^3\Phi^2\right)
\end{aligned} \tag{A5}
$$

where

$$
\begin{aligned}
\xi &\equiv \frac{T-\mu}{\sigma} \\
\phi &\equiv \phi(\xi) \equiv \frac{\exp\left(\left(\frac{-1}{2}\right)\xi^2\right)}{\sqrt{2\pi}} \\
\Phi &\equiv \Phi(\xi) \equiv \int_{-\infty}^{\xi} \phi(x)dx
\end{aligned}
$$

and where $\phi(\xi)$ and $\Phi(\xi)$ are the pdf and CDF of the standard normal distribution. Note that numbers are used to index the terms in equation (26), whereas here, where we know what the $p = 2$ parameters represent, we employ the parameter names. However, there is a strict correspondence between the two. For example, $\psi_{112} \equiv \psi_{\mu\mu\sigma}$, etc.

### A3. Generalizing to Multiple Predictor Variables

[75] The previous results assume a trivial regression model with $K = 0$ predictor variables. However, the results immediately generalize to the case of multiple predictor variables. If we define $\mu \equiv \sum_{i=0}^{m} \mathbf{X}_{Ci}\beta_i$, then the full set of partial derivatives can easily be computed by applying the chain rule: $\frac{\partial \ln(\ell)}{\partial\beta_i} = \left(\frac{\partial \ln(\ell)}{\partial\mu}\right)\cdot\left(\frac{\partial\mu}{\partial\beta_i}\right) = \left(\frac{\partial \ln(\ell)}{\partial\mu}\right)\cdot X_i$. Thus we need only compute terms with respect to $\mu$ and $\sigma$, as was done in the previous section, and the rest follow directly.

### A4. Bias and Variance of $\hat{\sigma}^2$

[76] The first-order bias and variance of $\hat{\sigma}^2$ can be computed by reparameterizing the model in terms of $\sigma^2$ instead of $\sigma$ and applying the SB formulas. There is, however, a simpler way. If we define $f(x) \equiv x^2$, then $f(\hat{\sigma})$ can be expanded in a Taylor series around $\sigma$, and

$$
\begin{aligned}
Var[\hat{\sigma}^2] &\approx f'(\sigma)^2 \cdot Var[\hat{\sigma}] \\
&= 4\sigma^2 \cdot Var[\hat{\sigma}] \\
Cov[\hat{\sigma}^2, \hat{\mu}] &\approx f'(\sigma) \cdot Cov[\hat{\sigma}, \hat{\mu}] \\
&= 2\sigma \cdot Cov[\hat{\sigma}, \hat{\mu}] \\
Bias_1[\hat{\sigma}^2] &\approx f'(\sigma) \cdot Bias_1[\hat{\sigma}] + Var[\hat{\sigma}] + O(N^{-2}) \\
&= 2\sigma * Bias_1[\hat{\sigma}] + Var[\hat{\sigma}]
\end{aligned} \tag{A6}
$$

### A5. Expected Value of $\exp(s_A^2)$

[77] Assuming $\nu s_A^2/\sigma^2$ is a $\chi_\nu^2$ random variable, the noncentral moments of $s_A^2$ are given by [*Johnson et al.*, 1994, p. 420, equation 18.8]

$$
E\left[\left(s_A^2\right)^k\right] = \sigma^{2k}\left\{(2/\nu)^k\frac{\Gamma(\nu/2+k)}{\Gamma(\nu/2)}\right\} = \sigma^{2k}/d_k \tag{A7}
$$

and therefore $E[d_k(s_A^2)^k] = \sigma^{2k}$.

[78] The $\{c_k\}$ appearing in equation (40) are the coefficients in the series expansion of $\exp(b_1\sigma + b_2\sigma^2)$ as a function in $\sigma$ arising from the identity

$$
\exp\left(b_1\sigma + b_2\sigma^2\right) = \sum\frac{\left(b_1\sigma + b_2\sigma^2\right)^k}{k!} = \sum c_k\sigma^k \tag{A8}
$$

[79] Inserting the definition of $c_k$ and the result from equation (A7) into the definition for $g_\nu(z, b_1, b_2)$ yields

$$
\begin{aligned}
E\left[g_\nu\left(s_A^2, b_1, b_2\right)\right] &= E\left[\sum_{k=0}^{\infty} c_k d_k s_A^k\right] \\
&= \sum_{k=0}^{\infty} c_k E\left[d_k(s_A 2)^{k/2}\right] \\
&= \sum_{k=0}^{\infty} c_k\sigma^k \\
&= \exp\left(b_1\sigma + b_2\sigma^2\right)
\end{aligned} \tag{A9}
$$

Note that the $g_m(t)$ function (equation (21)) is a special case of $g_\nu(z, b_1, b_2)$ for which

$$
g_m(t) = g_{\nu=m}(t, 0, 1) \tag{A10}
$$

In passing, it should be noted that there are some notational differences in the literature. *Finney*'s [1941] function, $g(t)$, is equivalent to $g_m(\frac{m}{m+1}t)$, and *Stuart and Ord* [1979] employ $s^2$ to denote the MLE $\hat{\sigma}^2$.

## Notation

- $\vec{\mathbf{A}}$ vector of coefficients on $\sigma$ in $\omega$.
- $\boldsymbol{\beta}$ vector of regression coefficients of log load model.
- $\boldsymbol{\beta}^\star$ vector of regression coefficients of log concentration model.
- $Bias_1[\hat{\theta}]$ first-order bias of parameter estimator $\hat{\theta}$.
- $\mathbf{B}$ standardized first-order bias.
- $\vec{\mathbf{B}}$ vector of standardized biases.
- $\vec{\mathbf{C}}$ matrix of standardized covariance of $\hat{\omega}$.
- $Cov_1[\hat{\theta}_i, \hat{\theta}_j]$ first-order covariance of parameter estimators $\hat{\theta}_i$ and $\hat{\theta}_j$.
- $C(t)$ instantaneous constituent concentration at time $t$.
- $\epsilon(t)$ difference between $Y$ and at time $t$.
- $F_{\boldsymbol{\beta},\sigma}()$ cumulative distribution function of normal distribution with mean $\mathbf{X}_i\boldsymbol{\beta}$ and variance $\sigma^2$.
- $f_{\boldsymbol{\beta},\sigma}()$ probability density function of normal distribution with mean $\mathbf{X}_i\boldsymbol{\beta}$ and variance $\sigma^2$.
- $K$ number of predictor variables in regression.

$K_u$ units conversion coefficient.
$\ell$, the likelihood function, the basis of MLE and AMLE estimation.
$\mathbf{L}$ annual constituent load.
$\hat{\boldsymbol{L}}$ estimate of expected value of load.
$\hat{L}_A$ adjusted maximum likelihood load estimator.
$\hat{L}_{Duan}$ Duan's "smearing" load estimator.
$\hat{L}_{JK}$ jackknife maximum likelihood load estimator.
$\hat{L}_{MVUE}$ minimum variance unbiased load estimator.
$\hat{L}_{QMLE}$ (quasi) maximum likelihood load estimator.
$\hat{L}_{RC}$ rating curve load estimator.
$L(t)$ instantaneous constituent load at time $t$.
$\mathrm{MSE}_P$ mean square error of prediction of load.
$\mathbf{M}$ expected value of $\hat{\omega}$.
$\boldsymbol{\mu_L}$ expected value of true load.
$\mu_{\mathbf{X}\hat{\omega}}$ mean of $\mathbf{X}\hat{\boldsymbol{\omega}}$.
$N$ number of observations used to calibrate regression model.
$N_e$ number of subintervals over which loads are to be computed.
$\nu$ degrees of freedom in error distribution.
$\hat{\omega}$ vector of coefficients in reparameterized regression model.
$p$ the total number of parameters in the model.
$\mathbf{Q}$ vector of water discharge.
$Q(t)$ instantaneous water discharge at time $t$.
$s^2$ ordinary least squares estimator for variance of residuals.
$s_A^2$ AMLE estimator for variance of residuals.
$\boldsymbol{\Sigma}$ variance-covariance of $\hat{\omega}$.
$\sigma^2$ residual variance.
$\sigma^2_{\mathbf{X}\hat{\omega}}$ variance of $\mathbf{X}\hat{\boldsymbol{\omega}}$.
$T_i$ logarithm of censoring threshold corresponding to ith observation in calibration data set.
$\vec{\mathbf{V}}$ matrix of first-order standardized covariance of $\hat{\boldsymbol{\beta}}$.
$\mathbf{V}$ standardized first-order covariance.
$\mathbf{X}(t)$ vector of predictor variables evaluated at time $t$.
$\vec{\mathbf{X}}_C$ matrix of predictor variables used to calibrate regression model.
$\xi_i$ standardized censoring threshold corresponding to ith observation in calibration data set.
$Y$ logarithm of load
$Y^\star$ logarithm of concentration.

[80] **Acknowledgments.** The author thanks Greg Baier, Lou DeLong, Ed Gilroy, Bob Hirsch, Bill Kirby, Greg Schwarz, and Linda Phillips, all at the U.S. Geological Survey, who participated in this project at times and contributed in countless ways to the research on which this paper is based. The author would also like to thank four anonymous reviewers whose considerable efforts greatly improved the presentation. Finally, the author would like to thank Kathleen Gohn for pointing out far too many errors of grammar, style, and substance.

## References

Aitchison, J., and J. A. C. Brown (1981), *The Lognormal Distribution With Special Reference to Its Uses in Economics*, 2nd ed., Cambridge Univ. Press, New York.

Bowman, K. O., and L. R. Shenton (1988), *Properties of Estimators for the Gamma Distribution*, Marcel Dekker, New York.

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

Childress, C. J. O., W. T. Foreman, B. F. Connor, and T. J. Maloney (1999), New reporting procedures based on long-term method detection levels and some considerations for interpreations of water quality data provided by the U.S. Geological Survey National Water Quality Laboratory, *U.S. Geol. Surv. Open File Rep., 99-193*.

Christensen, V. G., and A. C. Ziegler (2000), Regression analysis and real-time water-quality monitoring to estimate constituent concentrations, loads, and yields in the Little Arkansas River, south-central Kansas, 1995–99, *U.S. Geol. Surv. Water Resour. Invest. Tech. Rep., 2000-4126*.

Cohn, T. A. (1988), Adjusted maximum likelihood estimation of the moments of lognormal populations from type 1 censored samples, *U.S. Geol. Surv. Open File Rep., 88-350*.

Cohn, T. A. (1995), Recent advances in statistical methods for the estimation of sediment and nutrient transport in rivers, *U.S. Natl. Rep. Int. Union Geod. Geophys. 1991–1994, Rev. Geophys.*, *33*, 1117–1124.

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

Cohn, T. A., D. L. Caulder, E. J. Gilroy, L. D. Zynjuk, and R. M. Summers (1992a), The validity of a simple statistical model for estimating fluvial constituent loads: An empirical study involving nutrient loads entering Chesapeake Bay, *Water Resour. Res.*, *28*(9), 2353–2364.

Cohn, T. A., E. J. Gilroy, and W. G. Baier (1992b), Estimating fluvial transport of trace constituents using a regression model with data subject to censoring, paper presented at the Joint Statistical Meetings, Am. Stat. Assoc., Boston.

Colby, B. R. (1955), The relationship of sediment discharge to streamflow, technical report, U.S. Geol. Surv., Reston, Va.

Crawford, C. G. (1991), Estimation of suspended-sediment rating curves and mean suspended-sediment loads, *J. Hydrol.*, *129*, 331–348.

Crawford, C. G. (1996), Estimating mean constituent loads in rivers by the rating-curve and flow-duration, rating-curve methods, Ph.D. thesis, Indiana Univ., Bloomington.

David, H. A. (1981), *Order Statistics*, 2nd ed., 360 pp., John Wiley, Hoboken, N. J.

DeLong, L. L. (1982), Water quality of streams and springs, Green River Basin, Wyoming, *U.S. Geol. Surv. Open File Rep., 82-4008*.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977), Maximum likelihood from incomplete data via the EM algorithm, *J. R. Stat. Soc., Ser. B*, *39*, 1–22.

Draper, N. R., and H. Smith (1981), *Applied Regression Analysis*, 2nd ed., 709 pp., John Wiley, Hoboken, N. J.

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

Efron, B. (1982), The jackknife, the bootstrap and other resampling plans, *CBMS-NSF Monogr. 38*, Soc. for Ind. and Appl. Math., Philadelphia, Pa.

Ferguson, R. I. (1986), River loads underestimated by rating curves, *Water Resources Research*, *22*(1), 1997–2004.

Finney, D. J. (1941), On the distribution of a variate whose logarithm is normally distributed, *J. R. Stat. Soc. Suppl.*, *7*, 155–161.

Government Accountability Office (2003), Water quality: Improved EPA guidance and support can help states develop standards that better target cleanup efforts, *Tech. Rep. GAO-03-308*, Washington, D. C.

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

Gilroy, E. J., W. H. Kirby, T. Cohn, and G. D. Glysson (1990b), Discussion of "Uncertainty in suspended sediment transport curves" by Mcbean and Al-nassri, *J. Hydraul. Eng.*, *116*(1), 143–145.

Hosking, J. R. M. (1985), A correction for the bias of maximum likelihood estimators of Gumbel parameters—Comment, *J. Hydrol.*, *78*, 393–396.

Johnson, N. L., S. Kotz, N. Balakrishnan (1994), *Continuous Univariate Distributions*, 2nd ed., 756 pp., John Wiley, Hoboken, N. J.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T.-C. Lee (1985), *The Theory and Practice of Econometrics*, 1019 pp., John Wiley, Hoboken, N. J.

Kotz, S., N. Balakrishnan, and N. L. Johnson (2000), *Continuous Multivariate Distributions*, 2nd ed., 702 pp., John Wiley, Hoboken, N. J.

Likěs, J. (1980), Variance of the MVUE for lognormal variance, *Technometrics*, *22*(2), 253–258.

Loucks, D. P., J. R. Stedinger, and D. A. Haith (1981), *Water Resource Systems Planning and Analysis*, 559 pp., Prentice-Hall, Upper Saddle River, N. J.

Miller, C. R. (1951), Analysis of flow-duration, sediment-rating curve method of computing sediment yield, technical report, U.S. Bur. of Reclam., Washington, D. C.

Miller, D. M. (1984), Reducing transformation bias in curve fitting, *Am. Stat.*, *38*(2), 124–126.

National Research Council (2001), *Assessing the TMDL Approach to Water Quality Management*, 109 pp., Natl. Acad. Press, Washington, D. C.

Robinson, P. M. (1982), On the asymptotic properties of estimators of models containing limited dependent variables, *Econometrica*, *50*(1), 27–41.

Rohatgi, V. K. (1976), *An Intorduction to Probability and Theory and Mathematical Statistics*, 684 pp., John Wiley, Hoboken, N. J.

Runkel, R., C. G. Crawford, and T. A. Cohn (2004), Load estimator (LOADEST): A Fortran program for estimating constituent loads in streams and rivers, *U.S. Geol. Surv. Techniques Methods, Book 4, Chap. A5*.

Shenton, L. R., and K. O. Bowman (1977), *Maximum Likelihood Estimation in Small Samples*, 186 pp., Oxford Univ. Press, New York.

Stuart, A., and J. K. Ord (1979), *The Advanced Theory of Statistics*, vol. 2, *Inference and Relationship*, 4th ed., 748 pp., Oxford Univ. Press, New York.

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

Tobin, J. (1958), Estimation of relationships for limited dependent variables, *Econometrica*, *26*, 24–36.

van Zwet, W. R. (1966), Bias in estimation from type I censored samples, *Stat. Neer.*, *20*(1), 143–148.

Vogel, R. M., J. R. Stedinger, and R. P. Hooper (2003), Discharge indices for water quality loads, *Water Resour. Res.*, *39*(10), 1273, doi:10.1029/2002WR001872.

T. A. Cohn, U.S. Geological Survey, MS 415, Reston, VA 20192, USA. (tacohn@usgs.gov)