

**LOAD ESTIMATOR (LOADEST):
A FORTRAN PROGRAM FOR ESTIMATING
CONSTITUENT LOADS IN STREAMS AND RIVERS**

Techniques and Methods Book 4, Chapter A5

Load Estimator (LOADEST):

**A FORTRAN Program for Estimating
Constituent Loads in Streams and Rivers**

By Robert L. Runkel, Charles G. Crawford, and Timothy A. Cohn

Techniques and Methods Book 4, Chapter A5

U.S. Department of the Interior
Gale A. Norton, Secretary

U.S. Geological Survey
Charles G. Groat, Director

U.S. Geological Survey, Reston, Virginia: 2004

For more information about the USGS and its products:
Telephone: 1-888-ASK-USGS
World Wide Web: <http://www.usgs.gov/>

Any use of trade, product, or firm names in this publication is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this report is in the public domain, it contains copyrighted materials that are noted in the text. Permission to reproduce those items must be secured from the individual copyright owners.

Contents

Abstract	1
1 Introduction	2
1.1 Background	2
1.2 Related Reading	2
1.3 Report Organization	2
1.4 Acknowledgments	2
2 Theory	3
2.1 Background – Linear Regression Approach to Load Estimation	3
2.2 Load Estimation Methods used within LOADEST	4
2.2.1 Maximum Likelihood Estimation (MLE)	5
2.2.2 Adjusted Maximum Likelihood Estimation (AMLE)	5
2.2.3 Least Absolute Deviation (LAD)	6
2.2.4 Summary of MLE, AMLE, and LAD for Load Estimation	6
2.3 Multicollinearity and Centering	7
2.4 Model Selection	7
3 User’s Guide	8
3.1 Input/Output Structure	8
3.2 Input Files	9
3.2.1 The Control File	9
3.2.2 The Header File	9
3.2.3 The Calibration File	15
3.2.4 The Estimation File	17
3.3 LOADEST Execution	18
3.3.1 Execution under Unix	18
3.3.2 Execution under Windows	18
3.4 Output Files	18
3.4.1 The echo.out file	19
3.4.2 Constituent Output Files	19
3.4.3 Residual Output Files	19
3.4.4 Individual Load Files	20
4 LOADEST Applications	21
4.1 Application 1: Analysis of an Uncensored Constituent using a Predefined Model	21
4.2 Application 2: Analysis of an Uncensored Constituent using a Seasonal Model	29
4.3 Application 3: Analysis of a Censored Constituent using a Seasonal Model	32
4.4 Application 4: Multiple Constituents with Automated Model Selection	35
4.5 Application 5: User-defined Model with an Additional Data Variable	40
4.6 Application 6: Regression Model for Concentration	45

5 Software Guide	49
5.1 Supported Platforms	49
5.2 Software Distribution	49
5.3 Installation	50
5.4 Compilation	52
5.5 Software Overview	52
5.5.1 Software Development	52
5.5.2 Include Files	53
5.5.3 Fatal Errors and Warnings	54
References	55
Appendix 1. Assignment of Detection Limits.....	57
Appendix 2. Example echo.out File from Application 4.....	61
Appendix 3. Example Constituent Output File from Application 4.....	66

Figures

1. LOADEST Input/Output files.....	8
2. Example residual output file.....	20
3. Control file for Application 1.....	22
4. Header file for Application 1.....	23
5. Partial listing of the calibration file for Application 1.....	24
6. Partial listing of the estimation file for Application 1.....	25
7. Calibrated regression model for the Illinois River.....	26
8. Estimates of log load versus model residuals from AMLE regression, Illinois River.....	27
9. Normal probability plot for the model residuals of the AMLE regression, Illinois River.....	28
10. Header file for Application 2.....	30
11. Calibrated regression model for the St. Joseph River.....	31
12. Seasonal AMLE load estimates for the St. Joseph River.....	32
13. Partial listing of the calibration file for Application 3.....	33
14. Monthly AMLE load estimates for the Potomac River.....	35
15. Header file for Application 4.....	36
16. Calibration file for Application 4.....	38
17. Residual plot for model number 7, White River IN.....	39
18. Histogram of estimated AMLE loads, White River IN.....	40
19. Header file for Application 5.....	42
20. Partial listing of the calibration file for Application 5.....	43
21. Partial listing of the estimation file for Application 5.....	44
22. Daily AMLE load estimates for the Little Arkansas River.....	45
23. Header file for Application 6.....	47
24. Partial listing of the calibration file for Application 6.....	48
25. Partial listing of the estimation file for Application 6.....	48
26. LOADEST directory structure.....	51

Tables

1. The control file	9
2. The header file - record types 1-2, title and print option	10
3. The header file - record type 3, standard error option	10
4. The header file - record type 4, load option	11
5. The header file - record type 5, number of user-defined seasons	11
6. The header file - record type 6, season begin and end dates	12
7. The header file - record type 7, model number	12
8. The header file - record type 8, begin and end months	13
9. The header file - record types 9 and 10, number of variables	13
10. The header file - record type 11, field 1, data variable name	14
11. The header file - record type 11, field 2, transformation	14
12. The header file - record type 12, number of constituents	15
13. The header file - record type 13, constituent specific variables	15
14. The calibration file - record type 1, calibration data set	16
15. Constituent concentrations in the calibration file: possible states	16
16. The estimation file - record type 1, number observations per day	17
17. The estimation file - record type 2, time series of data variables	17
18. Calibration results, Application 1	26
19. Mean load estimates, Application 1	28
20. AMLE model coefficients, Application 3	34
21. Correlation between explanatory variables, AMLE regression, Application 3	34
22. Supported systems	49
23. Files to download for Unix and Windows systems	50
24. Development environments	53
25. Maximum dimensions and default values from <code>fmodules.inc</code>	53
26. Sensitivity of AMLE to value of <code>dlim</code>	57
27. Assignment of <code>dlim</code> , Case #2	58
28. Assignment of <code>dlim</code> , Case #3	59
29. Assignment of <code>dlim</code> , Case #3 reordered	60

Load Estimator (LOADEST): A FORTRAN Program for Estimating Constituent Loads in Streams and Rivers

By Robert L. Runkel, Charles G. Crawford, and Timothy A. Cohn

Abstract

LOAD ESTimator (LOADEST) is a FORTRAN program for estimating constituent loads in streams and rivers. Given a time series of streamflow, additional data variables, and constituent concentration, LOADEST assists the user in developing a regression model for the estimation of constituent load (calibration). Explanatory variables within the regression model include various functions of streamflow, decimal time, and additional user-specified data variables. The formulated regression model then is used to estimate loads over a user-specified time interval (estimation). Mean load estimates, standard errors, and 95 percent confidence intervals are developed on a monthly and(or) seasonal basis.

The calibration and estimation procedures within LOADEST are based on three statistical estimation methods. The first two methods, Adjusted Maximum Likelihood Estimation (AMLE) and Maximum Likelihood Estimation (MLE), are appropriate when the calibration model errors (residuals) are normally distributed. Of the two, AMLE is the method of choice when the calibration data set (time series of streamflow, additional data variables, and concentration) contains censored data. The third method, Least Absolute Deviation (LAD), is an alternative to maximum likelihood estimation when the residuals are not normally distributed. LOADEST output includes diagnostic tests and warnings to assist the user in determining the appropriate estimation method and in interpreting the estimated loads.

This report describes the development and application of LOADEST. Sections of the report describe estimation theory, input/output specifications, sample applications, and installation instructions.

1 INTRODUCTION

1.1 Background

This report describes the development and use of LOAD ESTimator (LOADEST), a FORTRAN program for estimating constituent loads in streams and rivers. LOADEST is based on two previously undocumented software programs known unofficially as LOADEST2 and ESTIMATOR [see Crawford (1996) and Cohn (1988) for relevant details].

1.2 Related Reading

Successful application of LOADEST requires considerable knowledge of statistics, multiple linear regression, and load estimation. Although an overview of load estimation theory is provided in Section 2, the overview is intentionally brief as the underlying statistical concepts are described in detail elsewhere. Software users should therefore review the relevant concepts and techniques presented in statistical texts (for example: Judge and others, 1988; Draper and Smith, 1998; Helsel and Hirsch, 2002).

1.3 Report Organization

The remaining sections of this report are as follows. Section 2 provides a description of the theoretical constructs underlying the problem of load estimation. This section includes descriptions of the load estimation problem and the equations used for load estimation. Section 3, a User's Guide, presents the input and output requirements of the FORTRAN computer program. This section describes the format of the input/output files and various user options. Section 4 presents several applications of LOADEST and includes example input and output files. The final section, a Software Guide (Section 5), describes how to obtain LOADEST, installation procedures, and several programming features.

1.4 Acknowledgments

Many of the concepts and techniques underlying LOADEST are based on earlier work by Tim Cohn and Charlie Crawford of the U.S. Geological Survey (USGS). This project was initiated following a workshop organized by Bob Broshears and Dan Hippe of the USGS. Other workshop participants included Doug Glysson, Dennis Helsel, Dave Lorenz, Dave Mueller, Skip Vecchia, and Andy Ziegler. The authors wish to thank Dan Hippe, Michael Langland, Dave Lorenz, and Greg Schwarz for their review comments, Jeff Deacon and Norm Spahr for additional testing and commentary, and Alene Brogan, Betty Palcsak, and Loretta Ulibarri for editorial assistance and report formatting.

2 THEORY

2.1 Background – Linear Regression Approach to Load Estimation

It is often desirable to know the mass of suspended sediment or chemical constituent entering a lake or estuary from a river or stream. In general, total mass loading over an arbitrary time period, τ , is given by:

$$L_{\tau} = \int_0^{\tau} QC dt \quad (1)$$

where C is concentration [M/L^3], L_{τ} is total load [M], Q is instantaneous streamflow [L^3/T], and t is time [T]¹. Direct use of equation 1 is problematic as continuous estimates of Q and C are rarely available. Further, although discrete values of Q are readily available for many locations (streamflow estimates from USGS gaging stations, for example), values of C are considerably less common due to the expense of sample collection and analysis. Load estimates are therefore more commonly given by:

$$\hat{L}_{\tau} = \Delta t \sum_{i=1}^{NP} (\hat{Q}C)_i = \Delta t \sum_{i=1}^{NP} \hat{L}_i \quad (2)$$

where \hat{L}_i [$=\hat{Q}C_i$] is an estimate of instantaneous load [M/T], \hat{L}_{τ} is an estimate of total load [M], NP is the number of discrete points in time, and Δt is the time interval represented by the instantaneous load [T]. Mean load for time period τ is then given by

$$\bar{L} = \frac{\hat{L}_{\tau}}{\Delta t NP} \quad (3)$$

where \bar{L} is the mean load [M/T]. Calculation of loads using equations 2 and 3 is contingent upon two assumptions. First, each estimate of instantaneous load [$\hat{L}_i = \hat{Q}C_i$] is assumed to represent the mass load over the discrete time interval (Δt). Second, the discrete time interval is constant; all of the NP discrete points in time will have the same Δt . For example, a common application is to calculate the mean load for a calendar year, by using daily estimates of streamflow. Under this application, there will be 365 discrete points in time ($NP=365$) with a time interval of one day ($\Delta t=1$ day). Each estimate of instantaneous load represents average conditions for a given day.

As described by Cohn (1995), several techniques are available for estimating total load, \hat{L}_{τ} . Of these techniques, one based on linear regression is used within LOADEST. In its simplest form, the regression approach proceeds as follows. First, a linear model is formed in which the log of instantaneous load is related to one or more explanatory variables:

$$\ln(\hat{L}) = a_0 + \sum_{j=1}^{NV} a_j X_j \quad (4)$$

¹The fundamental units of Mass [M], Length [L], and Time [T] are used throughout this section. Specific units are introduced in Section 3.

where a_0 and a_j are model coefficients, NV is the number of explanatory variables, and X_j is an explanatory variable². Equation (4) is then exponentiated to yield an estimate of instantaneous load:

$$\hat{L}_{RC} = \exp \left(a_0 + \sum_{j=1}^M a_j X_j \right) \quad (5)$$

where \hat{L}_{RC} is a “rating curve” estimate of instantaneous load. Development of load estimates using equations 4 and 5 is thus a 3-step process:

(1) **Model Formulation.** The form of the linear model (the right-hand side of equation 4) is determined based on the user’s knowledge of the hydrologic and biogeochemical system. Each explanatory variable (X_j) is a function of a data variable (streamflow or time, for example) that is thought to influence instantaneous load. The number and form of explanatory variables is highly dependent on the system under study and the constituent of interest. A simple model with a single explanatory variable (log streamflow) is often sufficient for prediction of suspended-sediment load (Crawford, 1991), whereas a model with six explanatory variables based on various functions of streamflow and time is often applicable to nutrients (Cohn and others, 1992a). Additional guidance on model formulation is provided elsewhere (Judge and others, 1988; Draper and Smith, 1998; Helsel and Hirsch, 2002).

(2) **Model Calibration.** Given the form of the regression model, a time series of constituent load and the explanatory variables is used to develop the model coefficients (a_0 and a_j , equ. 4) by using ordinary least squares (OLS) regression. The regression equation then is used to calculate estimates of log load [$\ln(\hat{L})$] for each observation in the time series (the calibration data set). Residual error for each observation is equal to the difference between observed and estimated values of log load [$\ln(L) - \ln(\hat{L})$].

(3) **Load Estimation.** Estimates of the instantaneous load are obtained using the retransformed version of the regression model (equ. 5) and a time series of explanatory variables (the estimation data set). Individual estimates of instantaneous load then are used to determine the total (equ. 2) or mean (equ. 3) load.

As outlined above, estimation of constituent loads using the regression approach is theoretically straightforward. Several statistical complications arise, however, when dealing with real-world data. Load calculations within LOADEST are therefore more complex than the calculations described above. Three of these complicating factors (retransformation bias, data censoring, and nonnormality) are described below, where the three load estimation methods used within LOADEST are detailed. Additional issues that are germane to all three methods are described in Sections 2.3 and 2.4.

2.2 Load Estimation Methods used within LOADEST

The load estimation process is complicated by retransformation bias, data censoring, and nonnormality. As noted by Ferguson (1986), rating curve estimates (equ. 5) of instantaneous load are biased; estimates may underestimate the true load by as much as 50 percent. This retransformation bias is addressed by introducing bias correction factors for the calculation of instantaneous load. Data censoring occurs when one or more observations used in the calibration step have constituent concentrations that are less than the laboratory detection limit (Gilbert, 1987). Although substitution (setting C equal to one-half the detection limit, for example) appears to be a simple remedy for the replacement of less-than values, none of the substitution methods commonly used yield adequate results (Helsel and Cohn, 1988). A more rigorous treatment of censored data is therefore required. A final complication is the assumption of OLS regression that the model residuals are normally distributed. Alternate methods for estimating model coefficients are applicable when model residuals do not follow a normal

²The i subscript is omitted from L in equation 4 and all subsequent equations.

distribution. Because of these complications, LOADEST provides three methods for load estimation; each method is described below.

2.2.1 Maximum Likelihood Estimation (MLE)

As an alternative to OLS regression, model coefficients (a_0 and a_j , equ. 4) may be calculated using the method of maximum likelihood (MLE). When the calibration data set includes censored data, implementation of MLE also is known as tobit regression (Helsel and Hirsch, 2002). As with OLS, tobit regression assumes that model residuals are normally distributed with constant variance.

Given the model coefficients provided by regression, estimates of instantaneous load may be obtained by retransforming equation 4. When the calibration data set is uncensored, the bias correction factor of Bradu and Mundlak (1970) provides a minimum variance unbiased estimate (MVUE) of instantaneous load (Cohn and others, 1989):

$$\hat{L}_{MVUE} = \exp \left(a_0 + \sum_{j=1}^M a_j X_j \right) g_m(m, s^2, V) \quad (6)$$

where \hat{L}_{MVUE} is the MLE estimate of instantaneous load, m is the number of degrees of freedom, s^2 is the residual variance, and V is a function of the explanatory variables (Cohn and others, 1989). The model coefficients in equation 6 (a_0 and a_j) are estimated by maximum likelihood; the bias correction factor [$g_m(m, s^2, V)$] is an approximation of the infinite series given in Finney (1941). Within LOADEST, $g_m(m, s^2, V)$ is replaced by a similar function, phi (Likes, 1980).

Under the MLE method, estimates of instantaneous load are developed for all of the observations in the estimation data set using equation 6. Mean load estimates for various time periods then are calculated using equation 3 (where $\hat{L}_\tau = \hat{L}_{MVUE}$). Standard errors reflecting the uncertainty in each estimate of mean load are calculated by using the method described by Likes (1980) and Gilroy and others (1990) (for specifics, see equations 9–25 in Gilroy and others, 1990).

2.2.2 Adjusted Maximum Likelihood Estimation (AMLE)

For the case of censored data, model coefficients estimated by tobit regression (MLE, Section 2.2.1) exhibit first-order bias. In addition, the Bradu-Mundlak bias correction factor (g_m , equation 6) results in biased estimates of instantaneous load. By using adjusted maximum likelihood estimation (AMLE, Cohn 1988; Cohn and others, 1992b), first order bias in the model coefficients is eliminated using the calculations given in Shenton and Bowman (1977). A “nearly unbiased” (Cohn 1988) estimate of instantaneous load then is given by:

$$\hat{L}_{AMLE} = \exp \left(a_0 + \sum_{j=1}^M a_j X_j \right) H(a, b, s^2, \alpha, \kappa) \quad (7)$$

where \hat{L}_{AMLE} is the AMLE estimate of instantaneous load, a and b are functions of the explanatory variables (Cohn and others, 1992b), α and κ are parameters of the gamma distribution, and s^2 is the residual variance. The model coefficients in equation 7 (a_0 and a_j) are maximum likelihood estimates corrected for first-order bias; the bias correction factor [$H(a, b, s^2, \alpha, \kappa)$] is an approximation of the infinite series given in Cohn and others (1992b).

Under AMLE, estimates of instantaneous load are developed for all of the observations in the estimation data set using equation 7. Mean load estimates for various time periods then are calculated using equation 3 (where $\hat{L}_\tau = \hat{L}_{AMLE}$). The uncertainty associated with each estimate of mean load is expressed in terms of the standard error (SE) and the standard error of prediction (SEP). The SE for each mean load estimate (Cohn and others, 1992b; equ. 35) represents the variability that may

be attributed to the model calibration (parameter uncertainty). Calculation of the SEP begins with an estimate of parameter uncertainty (the SE) and adds the unexplained variability about the model (random error). Because SEP incorporates parameter uncertainty and random error, it is larger than SE and provides a better description of how closely estimated loads correspond to actual loads. The SEP is therefore the preferred method of describing uncertainty in loads and is used within LOADEST to develop 95 percent confidence intervals for each estimate of mean load.

2.2.3 Least Absolute Deviation (LAD)

All of the regression methods discussed thus far (OLS, MLE, AMLE) assume the model residuals are normally distributed with constant variance. When model residuals do not conform to the assumption, alternate techniques may be appropriate. One such technique, the least absolute deviation (LAD) method, is implemented within LOADEST. Model coefficients for LAD are developed using the regression method of Powell (1984), as implemented by Buchinsky (1994). Given the model coefficients, estimates of instantaneous load are developed using the “smearing” approach of Duan (1983):

$$\hat{L}_{LAD} = \exp\left(a_0 + \sum_{j=1}^M a_j X_j\right) \frac{\sum_{k=1}^n \exp(e_k)}{n} \quad (8)$$

where \hat{L}_{LAD} is the LAD estimate of instantaneous load, a_0 and a_j are model coefficients developed by the LAD regression, e is the residual error, and n is the number of uncensored observations in the calibration data set³.

LAD estimates of instantaneous load are developed for all of the observations in the estimation data set using equation 8. Mean load estimates for various time periods then are calculated using equation 3 (where $\hat{L}_\tau = \hat{L}_{LAD}$). Standard errors reflecting the uncertainty in each estimate of mean load are calculated using the jackknife method described by Efron (1982).

2.2.4 Summary of MLE, AMLE, and LAD for Load Estimation

The primary load estimation method used within LOADEST is AMLE. AMLE has been shown to have negligible bias when the calibration data set is censored (Cohn and others, 1992b). For the special case where the calibration data set is uncensored, the AMLE method converges to MLE (Cohn and others, 1992b), resulting in a minimum variance unbiased estimate of constituent loads. MLE estimates are provided as a check on AMLE results and as a means of comparing LOADEST results with standard statistical packages that implement MLE.

AMLE and MLE results are contingent upon the assumption that model residuals are normally distributed. Following model formulation and calibration (Section 2.1), AMLE residuals should be examined to see if the normality assumption is valid. Checks for normality include calculation of the PPCC (probability plot correlation coefficient; Vogel, 1986) and Turnbull-Weiss likelihood ratio (Turnbull and Weiss, 1978) statistics, construction of a normal-probability plot (the graphical analog of the PPCC; Helsel and Hirsch, 2002), and examination of standardized residuals. If the residuals do not adhere to the assumption of normality, AMLE (and MLE) results for censored data may not be optimal. Load estimates from the LAD method should therefore be considered in lieu of AMLE, as the LAD load estimates are not dependent on the normality assumption.

³Because of a lack of published results for the verification of the numerical algorithm, the LOADEST implementation of LAD is limited to the case of uncensored data. LAD results are omitted when the calibration data set contains censored observations.

2.3 Multicollinearity and Centering

Multicollinearity arises when one of the explanatory variables (equation 4) is related to one or more of the other explanatory variables (Helsel and Hirsch, 2002). The presence of collinear explanatory variables is undesirable because it confounds interpretation of model coefficients and tests of their significance. Causes of multicollinearity include natural phenomena, such as a positive relation between explanatory variables based on streamflow and precipitation, as well as mathematical artifacts, when one explanatory variable is a function of another explanatory variable. This latter cause is common in load estimation problems, when quadratic terms based on decimal time or log streamflow are included in the regression model (see, for example, the predefined models described in Section 3.2.2). In such a case, explanatory variables may be centered to eliminate the collinearity. The center of the calibration data, \tilde{T} , is given by (Cohn and others, 1992a):

$$\tilde{T} = \bar{T} + \frac{\sum_{k=1}^N (T - \bar{T})^3}{2 \sum_{k=1}^N (T - \bar{T})^2} \quad (9)$$

where N is the number of observations in the calibration data set, \bar{T} is the mean of the data, and T is the quantity to be centered (decimal time or log streamflow). Within LOADEST, \tilde{T} is subtracted from T , and the resulting “centered” values are used to develop the linear (decimal time, log streamflow) and quadratic (decimal time squared, log streamflow squared) explanatory variables. As a result, the linear and quadratic terms are orthogonal and no longer collinear.

2.4 Model Selection

As discussed in Section 3.2.2, LOADEST includes several predefined models that specify the form of the regression equation (the right-hand side of equ. 4). These models may be selected by the user based on the user’s knowledge of the hydrologic and biogeochemical system (“Model Formulation”, Section 2.1). Alternatively, the software provides an automated model selection option that selects the “best” model from the set of predefined models. Under this option, AMLE is used to determine model coefficients and estimates of log load (equ. 4; “Model Calibration”, Section 2.1) for each predefined model. Two statistics, the Akaike Information Criterion (AIC) and the Schwarz Posterior Probability Criterion (SPPC), then are computed for the calibrated model (Judge and others, 1988). The predefined model with the lowest value of the AIC statistic then is selected for use in load estimation (Section 2.1) (values of SPPC are provided for comparative purposes only and are not used directly in the model selection process).

3 USER'S GUIDE

This section, the LOADEST User's Guide, provides information on the use of LOADEST. Section 3.1 begins with a general description of the LOADEST input/output structure. Section 3.2 provides a detailed description of the required input files; model execution and the output file formats are described in Sections 3.3 and 3.4.

3.1 Input/Output Structure

This section describes the input and output files associated with LOADEST. A brief description of each file is presented here; more detailed descriptions are provided in Sections 3.2–3.4. The input/output structure of LOADEST is depicted in figure 1. The first input file, the control file, is used to specify the filenames of the remaining input files. Unlike the other input files, the filename for the control file is set to `control.inp` within the software, and its name cannot be changed. A second input file, the header file, defines the constituents of interest, specifies the form of the regression model, and sets load estimation options. A third input file, the calibration file, contains the information (concentration and data variables) that is used to calibrate the regression model (“Model Calibration,” Section 2.1). The final input file, the estimation file, contains the data variables for which load estimates are developed (“Load Estimation,” Section 2.1). The filenames of the header, calibration, and estimation files are specified by the user in the control file, `control.inp`

Input/Output Files

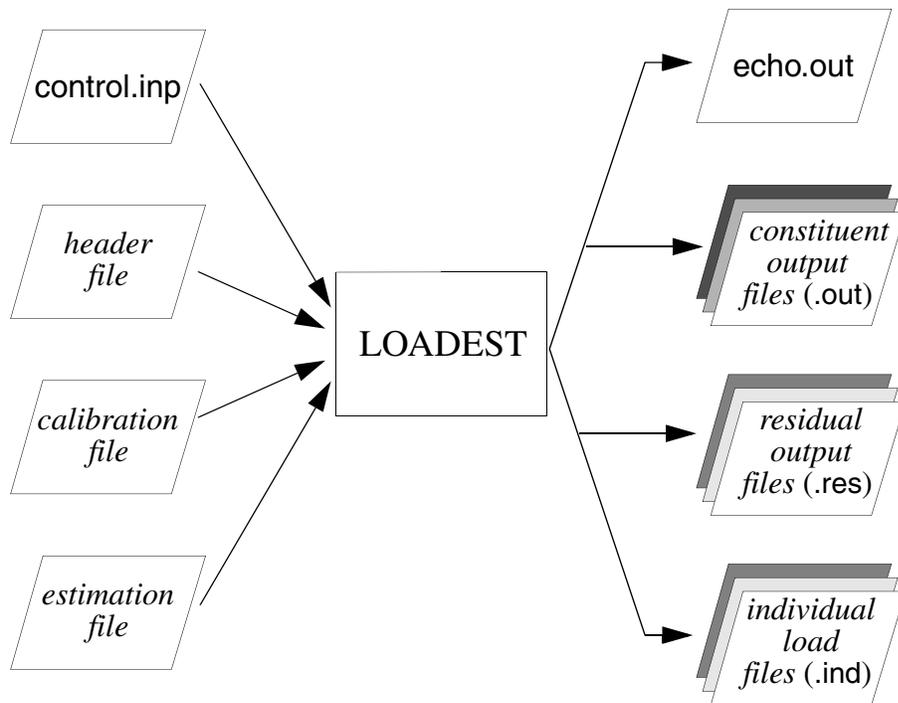


Figure 1. LOADEST Input/Output files.

Also shown in figure 1 are the output files created by LOADEST. Upon completion of a LOADEST run, the file `echo.out` contains an “echo” of the user-specified information (Section 3.4.1). In addition to the echo file, the software creates a constituent output file, residual output file, and individual load file for each constituent. As described in Section 3.4.2, the constituent output files contain a summary of the calibration and load estimation calculations. The residual output files contain the residuals from the AMLE regression and other quantities used to create residual plots (Section 3.4.3). The individual load

file contains the load estimates corresponding to the entries in the estimation input file (Section 3.4.4). Filenames for the constituent, residual, and individual load output files are created by the software by appending “.out”, “.res”, and “.ind” extensions to the user-specified constituent name (see Section 3.2.2).

3.2 Input Files

As described in Section 3.1, several input files must be assembled prior to LOADEST execution. Specific details relevant to each input file are provided in Sections 3.2.1–3.2.4. Details germane to all input files are briefly described below. Sample input files are presented in Section 4.

Record Types and Fields. In Sections 3.2.1–3.2.4, each input file is described in terms of a set of record types. Within each record type, one or more fields are used to specify various input parameters. In general, record types refer to lines in the input file, and fields correspond to specific columns within each record. In most cases, the specific columns for a given field are specified in the table that describes each record type. Some exceptions to the column requirements are noted in the text, when a given record type is in a free format, and placement of input in specific columns is not required. In addition to column specifications, each table specifies the field formats. A given field may be double precision (D), integer (I), or character (C). Double precision fields require entry of a numeric value that includes a decimal point; integer fields require entry of a numeric value without a decimal point. Alphanumeric values may be placed in character fields.

Internal Comments. LOADEST users may document their work by placing comments within any of the input files described below. All lines with a pound sign (#) placed in column number 1 will be treated as comments by the software. This feature is illustrated in the sample input files described in Section 4.

Common Pitfalls. The presence of tab characters within an input file may result in errors during program execution. LOADEST users should therefore avoid the use of tab characters when preparing the input files; placement of input values in the correct columns should be accomplished by inserting spaces rather than tab characters. Another common problem is the presence of blank lines within the input files. During program execution, an error will occur when the software attempts to read input values from a blank line (users who desire extra white space in their input files may specify internal comments as described above). Blank or extraneous lines placed at the end of the calibration and estimation files (Sections 3.2.3–3.2.4) also will also cause problems during program execution.

3.2.1 The Control File

The control file, control.inp, is used to specify the filenames for the various input files. The format of control.inp is shown in table 1, where the control file consists of three record types. Record type 1 specifies the filename for the header file, record type 2 specifies the name of the calibration file, and record type 3 specifies the name of the estimation file. An example control file is given in Section 4.1.

Table 1. The control file.
[C, Character]

Record type	Format	Column	Description
1	C	1-40	Filename for the <i>Header File</i>
2	C	1-40	Filename for the <i>Calibration File</i>
3	C	1-40	Filename for the <i>Estimation File</i>

3.2.2 The Header File

The header file is used to define the constituents of interest, specify the form of the regression model, and set load estimation options. The format of the header file is given in tables 2–13. The header file is created using the 13 record types discussed below.

Record Type 1 - Run Title (TITLE, table 2)

The first record in the header file is a title of up to 80 characters. This title is printed as part of the echo.out and the constituent output files.

Record Type 2 - Estimated Values Print Option (PRTOPT, table 2)

Mean loads and the corresponding standard errors are summarized in the constituent output files. Individual load estimates on which these summaries are based may be output at the user's request by using the estimated values print option, set in record type 2. If the print option is set to 0, individual load estimates are not included as part of the output; if the print option is set to 1, individual load estimates are included in a separate output file (Section 3.4.4).

Table 2. The header file - record types 1–2, title and print option.
[C, Character; I, Integer]

Record type	Input variable	Format	Column	Description
1	TITLE	C	1-80	Run title
2	PRTOPT	I	1-5	Estimated values print option

Record Type 3 - Standard Error Option (SEOPT, table 3)

The standard error provides a measure of the uncertainty associated with each estimate of mean load. Estimates of the standard error for AMLE load estimates may be based on detailed statistical computations ("exact" standard error) or a linear approximation. Due to the computational demands of calculating the exact standard error, software users may wish to use the linear approximation method when performing initial loading analyses. Computation of the exact standard error then may be requested when a final loading model is determined. Users may select the appropriate method for calculating the standard error using the standard error option, SEOPT (table 3).

As shown in table 3, setting SEOPT to 1 results in the use of a linear approximation for the standard error on the AMLE load estimates. Standard errors for MLE and LAD load estimates are not computed when SEOPT equals 1. Setting SEOPT to 2 also results in the use of a linear approximation for the AMLE standard error; in addition, standard errors for the remaining load estimates (MLE and LAD) are computed. Setting SEOPT to 3 results in computation of the exact standard error for the AMLE load estimates; standard errors for the remaining load estimates (MLE and LAD) also are computed. Note that use of the linear approximation for the AMLE standard error also affects the standard error of prediction and 95 percent confidence intervals for the AMLE loads (Section 2.2.2).

Software run time increases as the standard error option increases from 1 to 3 (approximate relative run time, table 3). Loading analyses conducted with SEOPT set to 2 will take approximately 10 times longer than analyses completed with SEOPT equal to 1. Loading analyses conducted with SEOPT set to 3 will take approximately 70 times longer than analyses completed with SEOPT equal to 1.

Table 3. The header file - record type 3, standard error option.
[I, Integer]

Input variable	Format	Column	Description
SEOPT	I	1-5	Standard error option

Specified value	Approximate relative run time	Standard error calculation
1	1	Calculates standard error for AMLE based linear approximation. Standard error is not calculated for MLE and LAD.
2	10	Calculates standard error for AMLE based linear approximation. Also calculate standard error for MLE and LAD.
3	70	Calculate exact standard error for AMLE. Also calculate standard error for MLE and LAD.

Record Type 4 - Load Option (LDOPT, table 7)

By default, LOADEST calculates the mean load for the entire estimation period (the entire length of time spanned by the estimation file, Section 3.2.4). Users also may request mean load estimates for seasonal and/or monthly time periods. The type of mean load calculations depends on the value of the load option, LDOPT (table 4). When LDOPT equals 0, only the default load calculations are completed (mean loads for the entire estimation period). Mean load calculations for a set of user-defined seasons are included when LDOPT equals 1, whereas load calculations for each month of data in the estimation file are included when LDOPT equals 2. Seasonal and monthly load calculations are both included when LDOPT equals 3. Users requesting seasonal output (LDOPT=1 or 3) also must define the seasonal time periods as described below (record types 5 and 6).

Table 4. The header file - record type 4, load option.

[I, Integer]

Input variable	Format	Column	Description
LDOPT	I	1-5	Load option

Specified value	Mean loads calculated for:
0	Entire estimation period
1	Entire estimation period; User-defined seasons
2	Entire estimation period; Each month of data in the estimation file
3	Entire estimation period; User-defined seasons; Each month of data in the estimation file

Record Types 5 and 6 - User-defined Seasons (optional)

Note: Record types 5 and 6 are omitted if the load option is not equal to 1 or 3 (LDOPT=0 or 2, record type 4).

Record types 5 and 6 are used to define seasonal time periods when seasonal output is requested (LDOPT=1 or 3). The number of user-defined seasons is first set using NSEAS (record type 5, table 5). Each season then is defined using record type 6 (table 6). Record type 6 contains two fields; each field has a free format such that placement of input in specific columns is not required. Field one specifies the season start date (SBEG), and field two specifies the season end date (SEND). Although the fields are in free format, values of SBEG and SEND must be in MMDD format, without embedded spaces (where MM equals the month and DD equals the day); the first two characters are used to specify the month (01–12) and the last two characters are used to specify the day (01–31). Record type 6 is used NSEAS times, once for each user-defined season.

Table 5. The header file - record type 5, number of user-defined seasons.

[I, Integer]

Input variable	Format	Column	Description
NSEAS	I	1–5	Number of user-defined seasons

Table 6. The header file - record type 6, season begin and end dates.

[I, Integer; Record type 6 is used once for each user-defined season (NSEAS times)]

Input variable	Format	Description
SBEG	I	Starting date for user-defined season (MMDD format)
SEND	I	Ending date for user-defined season (MMDD format)

Record Type 7 - Model Number (MODNO, table 7)

The next step in constructing the header file is to specify the form of the regression model. When specifying the regression model, the user has three options. First, the user may select one of the eleven predefined models by setting MODNO to the appropriate value (MODNO=1-11, table 7). Second, the user may select the automated model selection option by setting MODNO to 0. When the automated selection option is invoked (MODNO=0), the best regression model is selected from candidate regression models 1–9 based on the Akaike Information Criteria (Section 2.4). Regression models 10 and 11 are not included in the automated selection process. Third, the user may develop a user-defined regression model by setting MODNO equal to 99 and specifying the appropriate information in record types 9–11.

Table 7. The header file - record type 7, model number.

[I, Integer; $\ln Q = \ln(\text{streamflow})$ - center of $\ln(\text{streamflow})$; $dtime$ = decimal time - center of decimal time; per = period, 1 or 0 depending on defined period (record type 8)]

Input variable	Format	Column	Description
MODNO	I	1–5	Model number

Specified value	Regression model
0	automatically select best model from models 1-9.
1	$a_0 + a_1 \ln Q$
2	$a_0 + a_1 \ln Q + a_2 \ln Q^2$
3	$a_0 + a_1 \ln Q + a_2 dtime$
4	$a_0 + a_1 \ln Q + a_2 \sin(2\pi dtime) + a_3 \cos(2\pi dtime)$
5	$a_0 + a_1 \ln Q + a_2 \ln Q^2 + a_3 dtime$
6	$a_0 + a_1 \ln Q + a_2 \ln Q^2 + a_3 \sin(2\pi dtime) + a_4 \cos(2\pi dtime)$
7	$a_0 + a_1 \ln Q + a_2 \sin(2\pi dtime) + a_3 \cos(2\pi dtime) + a_4 dtime$
8	$a_0 + a_1 \ln Q + a_2 \ln Q^2 + a_3 \sin(2\pi dtime) + a_4 \cos(2\pi dtime) + a_5 dtime$
9	$a_0 + a_1 \ln Q + a_2 \ln Q^2 + a_3 \sin(2\pi dtime) + a_4 \cos(2\pi dtime) + a_5 dtime + a_6 dtime^2$
10	$a_0 + a_1 per + a_2 \ln Q + a_3 \ln Q per$
11	$a_0 + a_1 per + a_2 \ln Q + a_3 \ln Q per + a_4 \ln Q^2 + a_5 \ln Q^2 per$
99	user defined

Record Type 8 - Period Begin and End Months (optional)

Note: Record type 8 is omitted if the model number is not equal to 10 or 11 (MODNO<10 or MODNO=99, record type 7).

If model number 10 or 11 is selected, the user must define the interval over which the periodic terms (*per*, table 7) apply. This definition is accomplished using record type 8, where the begin and end months of the interval are specified. The first field in record type 8 specifies the beginning month of the interval, PBMON; the ending month of the interval, PEMON, is specified in the second field. PBMON is set equal to the numeric value of the first month (1=January, 2=February, 3=March, etcetera) during which the periodic terms apply, whereas PEMON is set equal to the numeric value of the last month during which the periodic terms apply.

Table 8. The header file - record type 8, begin and end months.

[I, Integer]

Input variable	Format	Column	Description
PBMON	I	1-5	Begin month for periodic terms
PEMON	I	6-10	End month for periodic terms

Record Types 9-11 - User-defined Model (optional)

Note: Record types 9-11 are omitted if the model number is not equal to 99 (MODNO<99, record type 7).

Record types 9-11 are used to specify the form of the regression model when a user-defined model is requested (MODNO=99). User-defined models consist of one or more explanatory variables ("Model Formulation," Section 2.1) that are formulated from the set of available data variables. Potential data variables include streamflow and adjusted decimal time (dtime = decimal time - center of decimal time), as with the formulation of explanatory variables for the predefined models (table 7). In addition, the user may input additional data variables (specific conductivity, for example) to be used in the formulation of explanatory variables. The number of additional data variables is given by NADDL in record type 9 (table 9). Note that NADDL is simply equal to zero if all the explanatory variables are based on streamflow and adjusted decimal time.

Table 9. The header file - record types 9 and 10, number of variables.

[I, Integer]

Record type	Input variable	Format	Column	Description
9	NADDL	I	1-5	Number of additional data variables
10	NEXPL	I	1-5	Number of explanatory variables

The number of explanatory variables in the user-defined regression model is given by NEXPL in record type 10 (table 9). Record type 11 then is used to specify the form of the NEXPL explanatory variables (tables 10 and 11). Field 1 of record type 11 identifies the name of the data variable (DVNAME, table 10), and field 2 specifies the transformation (TRANS, table 11) that is applied to the data variable to create the explanatory variable. Both fields have a free format such that placement of input in specific columns is not required. Valid values of DVNAME and TRANS are provided in tables 10 and 11. Record type 11 is used NEXPL times, once for each explanatory variable in the user-defined regression model.

Table 10. The header file - record type 11, field 1, data variable name.
[C, Character; Record type 11 is used once for each explanatory variable (NEXPL times)]

Input variable	Format	Description
DVNAME	C	Data variable name

Specified value	Data variable
Q	streamflow ¹
DTIME	adjusted decimal time (= decimal time - center of decimal)
ADDL <i>i</i>	additional data variable <i>i</i> , where $1 \leq i \leq \text{NADDL}$

¹adjusted streamflow [streamflow - center of ln(streamflow)] is used when the requested transformation (table 11) is LN or LNSQ

Table 11. The header file - record type 11, field 2, transformation.
[C, Character; Record type 11 is used once for each explanatory variable (NEXPL times)]

Input variable	Format	Description
TRANS	C	transformation applied to data variable

Specified value	Explanatory variable (transformation applied to data variable <i>x</i>)
NONE	x
SQ	x^2
SQRT	\sqrt{x}
LN	$\ln(x)$
LNSQ	$\ln(x)^2$
SIN2P	$\sin(2\pi x)$
SIN4P	$\sin(4\pi x)$
SIN6P	$\sin(6\pi x)$
COS2P	$\cos(2\pi x)$
COS4P	$\cos(4\pi x)$
COS6P	$\cos(6\pi x)$

Record Type 12 - Number of Constituents (NCONST, table 12)

The number of constituents that are to be analyzed, NCONST, is specified using record type 12. Calibration data for each of the NCONST constituents are entered in the calibration file as described in Section 3.2.3.

Table 12. The header file - record type 12, number of constituents.
[I, Integer]

Input variable	Format	Column	Description
NCONST	I	1-5	Number of constituents

Record Type 13 - Constituent Specific Variables (table 13)

Record type 13 sets three variables for each constituent (table 13). The first field in record type 13 is used to specify the constituent name, CNAME. CNAME is a text string up to 45 characters long. Because the value of CNAME is used to create the filename for the output files (Section 3.1), special characters should be avoided. The second and third fields are used to specify the concentration (UCFLAG) and load (ULFLAG) unit flags, respectively. If UCFLAG is set to 1, user supplied concentrations in the calibration file are in milligrams per liter; if UCFLAG is set to 2, user supplied concentrations are in micrograms per liter. The load unit flag is used to request load estimates in kilograms per day (ULFLAG=1), grams per day (ULFLAG=2), pounds per day (ULFLAG=3), or tons per day (ULFLAG=4). Because record type 13 is used for constituent-specific variables, it is used NCONST times, once for each constituent in the loading analysis.

Table 13. The header file - record type 13, constituent specific variables.
[C, Character; I, Integer; Record type 13 is used once for each constituent (NCONST times)]

Input variable	Format	Column	Units	Description
CNAME	C	1-45	---	Constituent name
UCFLAG	I	46-50	---	Concentration unit flag
ULFLAG	I	51-55	---	Load unit flag

3.2.3 The Calibration File

The information used to calibrate the regression model (observations of concentration and the data variables) is input in the calibration file. The calibration file consists of a single record type with five fields (table 14). Each field has a free format such that placement of input in specific columns is not required.

Table 14. The calibration file - record type 1, calibration data set.

[D, Double precision; I, Integer; Record type 1 is used once for each observation in the calibration file]

Input variable	Format	Units	Description
CDATE	I	---	Date of observation (YYYYMMDD format)
CTIME	I	---	Time of observation (HHMM format)
CFLOW	D	cubic feet per second	Streamflow at CTIME on CDATE
CADDL ¹	D	as appropriate for additional variable	Additional data variable at CTIME on CDATE
CCONC ²	D	as defined by UCFLAG (record type 13, header file)	Constituent concentration at CTIME on CDATE

¹CADDL repeats horizontally (a value of CADDL is specified for each of the NADDL variables).

²CCONC repeats horizontally (a value of CCONC is specified for each of the NCONST constituents).

As shown in table 14, the first three fields of record type 1 specify the date of the observation (CDATE), time of the observation (CTIME), and the streamflow (CFLOW). The fourth field, used to specify the values of additional data variables (CADDL), repeats horizontally; a value of CADDL is required for each of the NADDL data variables (record type 9, header file). When there are no additional data variables (NADDL=0), the CADDL field is omitted. The fifth field also repeats horizontally and is used to specify the concentration of each constituent (CCONC).

Although record type 1 is in free format, values of CDATE must be in YYYYMMDD format, without embedded spaces (where YYYY equals the year, MM equals the month, and DD equals the day); the first four characters are used to specify the year, characters 5 and 6 are used to specify the month, and characters 7 and 8 are used to specify the day (July 12, 2000, is specified as 20000712, for example). Similarly, values of CTIME must be in HHMM format without embedded spaces (where HH equals the hour in military time and MM equals the minute; 1:05 p.m. is specified as 1305, for example).

Specification of the CCONC field is dependent on the state of the observed data point. For a given constituent and observation, there are three possible states (table 15):

- (1) The observed constituent concentration is above the laboratory detection limit, and the data point is uncensored. When the data point is uncensored, CCONC is simply set equal to the observed constituent concentration.
- (2) The observed constituent concentration is below the laboratory detection limit, and the data point is censored. Censored data points are specified by using a less-than sign (<) as the first character in the CCONC field. The less-than sign is immediately followed by the laboratory detection limit for the constituent of interest (no spaces should appear between the less-than sign and the detection limit).
- (3) The constituent concentration is unavailable, and the data point is missing. Missing data points are specified by setting CCONC to zero or a negative number; when CCONC is less than or equal to zero, the observation will not be used during calibration.

Table 15. Constituent concentrations in the calibration file: possible states.

Possible state	Specified by setting CCONC equal to:
Uncensored observation	The observed constituent concentration
Censored observation	<XXX, where XXX is the laboratory detection limit
Missing observation	Zero or a negative number

When the calibration data set includes one or more censored observations, values of CCONC for censored observations are used to assign detection limits for uncensored observations. This assignment of detection limits is required for load estimation under AMLE. The general procedure for assigning detection limits is adequate in most cases; however, detection limit assignment can be problematic for the case of multiple detection limits (MDL). Users with MDL data that require highly accurate AMLE results should consult appendix 1.

Because multiple observations are required to calibrate the regression model, record type 1 is used multiple times (once for each observation). Model calibration within LOADEST requires 12 or more nonzero observations for a given constituent, 7 of which must be uncensored. Record type 1 is therefore used 12 or more times.

3.2.4 The Estimation File

The time series of data variables used to develop load estimates is specified in the estimation file. The format of the estimation file is given in tables 16–17. The estimation file is created using the two record types discussed below.

Record Type 1 - Number of Observations per day (NOBSPD, table 16)

The first record type is used to specify the number of observations of the data variables per day, NOBSPD (table 16). The number of observations per day must be less than or equal to 24.

Table 16. The estimation file - record type 1, number observations per day.

[I, Integer]

Input variable	Format	Column	Description
NOBSPD	I	1–5	Number of observations of the data variables per day

Record Type 2 - Time Series of Data Variables (table 17)

The second record type in the estimation file has four fields (table 17). Each field has a free format such that placement of input in specific columns is not required. The first three fields specify the date of the observation (EDATE), the time of the observation (ETIME), and the streamflow (EFLOW). As in the calibration file, the date (EDATE) and time (ETIME) must be in YYYYMMDD and HHMM format, respectively, without embedded spaces. The fourth field, used to specify the values of additional data variables (EADDL), repeats horizontally; a value of EADDL is required for each of the NADDL data variables (record type 9, header file). When there are no additional data variables (NADDL=0), the EADDL field is omitted.

Table 17. The estimation file - record type 2, time series of data variables.

[D, Double precision; I, Integer; Record type 2 is used once for each observation in the estimation file]

Input variable	Format	Units	Description
EDATE	I	---	Date of observation (YYYYMMDD format)
ETIME	I	---	Time of observation (HHMM format)
EFLOW	D	cubic feet per second	Streamflow at ETIME on EDATE
EADDL ¹	D	as appropriate for additional variable	Additional data variable at ETIME on EDATE

¹EADDL repeats horizontally (a value of EADDL is specified for each of the NADDL variables).

Because a time series of data variables is required for load estimation, record type 2 is used multiple times (once for each observation). Two requirements must be kept in mind when specifying the time series:

- When more than one observation per day is specified (NOBSPD > 1, record type 1), the times on a given day (ETIME) must be in ascending order.
- Each day must have the same number of observations (equal to NOBSPD).

3.3 LOADEST Execution

This section describes how to execute LOADEST under Unix and Windows operating systems⁴. Output files written during program execution are described in Section 3.4.

3.3.1 Execution under Unix

Before running LOADEST, the required input files (control, header, calibration, and estimation files) should be placed in the directory from which the software will be executed. To run LOADEST, enter the following command at the Unix prompt:

```
loadest
```

This command will initiate a LOADEST run, provided the loadest executable is in the current directory or within the user's path (see Section 5.3). During execution, run status information will be displayed on the screen indicating which constituent is currently being processed.

As an alternative to the command listed above, LOADEST may be run as a background process. When doing so, users may wish to redirect run status information that normally is displayed to the screen. Background runs are therefore best initiated by entering the following command at the Unix prompt:

```
loadest >& /dev/null &
```

3.3.2 Execution under Windows

Before running LOADEST, the required input files (control, header, calibration, and estimation files) should be placed in the folder from which the software will be executed. To run LOADEST, enter the following command at the MS-DOS prompt⁵:

```
loadest
```

This command will initiate a LOADEST run, provided the loadest executable is in the current directory or within the user's path (see Section 5.3). During execution, run status information will be displayed on the screen indicating which constituent currently is being processed.

LOADEST also may be executed by double clicking on the loadest executable that is displayed within Windows Explorer. This technique is not recommended, however, as run status information and LOADEST error messages displayed on the screen may not be visible following program execution.

3.4 Output Files

This section describes the output files written by LOADEST. The discussion includes descriptions of the echo (Section 3.4.1), constituent (Section 3.4.2), residual (Section 3.4.3), and individual load (Section 3.4.4) output files.

⁴The term "Unix" is used loosely herein to refer to both Linux and Unix-based operating systems. "Windows" as used herein refers to Microsoft Windows 98, Windows NT, Windows XP, Windows 2000, etcetera.

⁵Also known as the "Command Prompt" in Windows 2000 (see Programs/Accessories Menu).

3.4.1 The echo.out file

The output file `echo.out` contains an “echo” of user-specified information. This file also contains error messages generated during software execution. After completion of the LOADEST run, users should inspect `echo.out` to verify input values and to check for execution errors. An example `echo.out` file is provided in appendix 2.

3.4.2 Constituent Output Files

LOADEST creates a constituent output file for each constituent. The constituent output file is divided into two parts. Part Ia summarizes the calibration process (“Model Calibration,” Section 2.1), where regression models for constituent load are developed using the various methods (AMLE, MLE, and LAD). The corresponding AMLE regression model for constituent concentration is summarized in Part Ib. The calibration process within LOADEST uses constituent loads that are computed in units of kilograms per day. Regression models presented in Part Ia are therefore specific to these units. Users who desire regression equations corresponding to alternate load units may modify the intercept term, as follows:

$$a'_0 = a_0 + \ln(CF) \quad (10)$$

where a_0 is the intercept reported in Part Ia, a'_0 is the intercept term modified for the desired units, and CF is a conversion factor from kilograms per day to the desired units. A regression equation corresponding to load units of pounds per day, for example, can be developed using equation 10 with CF equal to 2.2046 (kilograms/day * 2.2046 pounds/kilogram = pounds/day). The remaining regression model coefficients (a_1, a_2 , etcetera) are unaffected by the choice of units. Note that the conversion shown in equation 10 is not required to obtain load estimates in the desired units (see description of Part II, below).

Part II of the constituent output file summarizes the mean load estimates provided by the various methods. Load estimates listed in Part II are in terms of the user-defined units (ULFLAG, Section 3.2.2). An example constituent output file is provided in appendix 3.

3.4.3 Residual Output Files

In addition to the constituent output files, LOADEST creates a residual output file for each constituent (Section 3.1). The following paragraphs describe the format of the residual output file, and several options for creating residual plots. In the discussion that follows, each line is described in terms of a number of fields; each field is 12 characters long.

The residual output file contains a line for each observation in the calibration file. As shown in figure 2, the first field of each line is the residual $[\ln(L) - \ln(\hat{L})]$, where L is in kilograms per day; Section 2.1]. The residual corresponds to the AMLE regression model; residuals for MLE and LAD are not output. The remaining four fields, from left to right, include the Z-score, the estimated value of log load $[\ln(\hat{L})]$, where L is in kilograms per day], the log of the observed streamflow (where streamflow is in cubic feet per second), and dtime (= decimal time - center of decimal time).

Residual Output File

1.76474E+00	1.71205E+00	-4.93522E+00	8.36404E+00	-1.00107E+00
4.76212E-01	3.90753E-01	-1.65618E+00	1.03546E+01	-9.19130E-01
-5.60472E-01	-2.75496E-01	-4.16801E+00	8.95803E+00	-8.28995E-01
9.06692E-01	7.08829E-01	-1.04118E+00	1.05891E+01	-7.68771E-01
-1.04916E+00	-9.38732E-01	-3.10358E+00	9.13884E+00	-6.67258E-01
-1.98977E+00	-1.51269E+00	-9.95524E-01	9.94750E+00	-5.57555E-01

↑ ↑ ↑ ↑ ↑

Residual Z-score Estimated Value ln(streamflow) dtime

Figure 2. Example residual output file (only first 6 lines are shown here).

The residual typically is plotted against one or more of the remaining four fields when conducting a residual analysis, an important part of the model formulation process (Section 2.1). A plot of model residuals versus their Z-scores, for example, yields a normal probability plot that may be used to assess the assumption of normality (Section 2.2.4; Helsel and Hirsch, 2002). Additional residual plots may be constructed to investigate the independence of model residuals and the assumption of constant error variance (Draper and Smith, 1998; Helsel and Hirsch, 2002).

Users have several options for creating residual plots. First, the residual output files may be imported into a spreadsheet as a space delimited file. The columns in the spreadsheet will then correspond to the fields described above (fig. 2). Another option is to use Grace, a free software package (<http://plasma-gate.weizmann.ac.il/Grace>). An example of this alternative is shown here; residual plots with the residual plotted on the y-axis are generated as follows⁶:

```
xmgrace -pexec 's0 symbol 2' -pexec 's0 line type 0' -block filename.res -bxy X:1
```

where `filename.res` is the name of the residual file and `X` is the field number of desired x variable. A residual plot with log streamflow on the x-axis, for example, would be created with `X=4`. Examples of Grace-generated residual plots are included in Section 4. Other options include Xgraph (Harrison, 1989) and numerous commercially-available software packages.

3.4.4 Individual Load Files

If the estimated values print option is set to 1 (PRTOPT=1, Section 3.2.2), individual load estimates are written to a separate output file with a `.ind` extension. The individual load file contains load estimates for each entry in the estimation input file. As in Part II of the constituent output file, load estimates are in the user-defined units (ULFLAG, Section 3.2.2).

⁶The `-free` and `-geometry` flags also may be of use; see Grace documentation.

4 LOADEST APPLICATIONS

In this section, several applications of LOADEST are presented. Each application is designed to illustrate one or more unique features of the software. The reader is encouraged to review all of the applications provided to obtain a complete overview of the software's capabilities.

Each application has four basic components: (a) a description of the features that are illustrated in the application, (b) a brief problem statement, (c) sample input files, and (d) results. The sample input files supplement the description of the input requirements presented in Section 3.2. These sample input files are annotated to illustrate how the record types are used within the various files. Due to space limitations, not all of the input files are shown for each application. A complete set of input files is available as part of the software distribution (Section 5.2). Note that the observed data used in the applications are for demonstration purposes only; individual data sets may not have been checked for accuracy and compliance with USGS quality-assurance guidelines. Also note that the description of results provided for each application is intentionally limited, and a complete description of the statistical aspects associated with a given problem is beyond the scope of this report.

4.1 Application 1: Analysis of an Uncensored Constituent using a Predefined Model

This application illustrates the format of the input files used within LOADEST. The application provides an analysis of a single, uncensored constituent. A predefined model that describes the linear relation between the log of constituent load and log streamflow is calibrated and used for load estimation.

The data used in this application were collected from the Illinois River at Marseilles, Illinois (p. 257, Helsel and Hirsch, 2002). Ninety-six water-quality samples for total phosphorus collected between November 1974 and April 1985 are used in conjunction with observed streamflow data to calibrate the regression model. The calibrated model then is compared with the results presented in Helsel and Hirsch (2002). As with the example provided by Helsel and Hirsch (2002), the streamflow data used for estimation is identical to that used for calibration; no attempt is made to estimate loads for other values of streamflow. This approach allows for comparison between load estimates and the "true" load of the calibration data set. More realistic applications which utilize estimation data from an expanded time period (daily observations of streamflow, for example) are presented in Sections 4.2–4.6.

Sample input files corresponding to the Illinois River analysis are shown below. Figure 3 depicts the control file, where the names of the header, calibration, and estimation files are specified. These filenames are specified using record types 1–3 (fig. 3). (Note that the "Record Type" text in figure 3 and the figures that follow are for illustrative purposes only and are not part of the actual input file).

Control File for Application 1

```
#####
#
# LOADEST control file
#
#
# line      name of the:
# ----      -
# 1         header file
# 2         calibration file
# 3         estimation file
#
#####
Record Type
1 header.inp
2 calib.inp
3 est.inp
```

Figure 3. Control file for Application 1.

Figure 4 depicts the header file for the Illinois River analysis. A descriptive title is specified using record type 1. The estimated values print option (PRTOPT, record type 2) is set to 1 such that load estimates corresponding to each observation in the estimation file are written to the individual output file. Computation of the exact standard error (SEOPT=3, record type 3) and the default load calculations (LDOPT=0, record type 4) are requested.

As discussed by Helsel and Hirsch (2002), exploratory data analysis and inspection of the residuals suggest a linear model that relates log load to log streamflow. This form of the regression model is provided by a predefined model within LOADEST:

$$\ln(\hat{L}) = a_0 + a_1 \ln Q \quad (11)$$

where the explanatory variable is centered (Section 2.3):

$$\ln Q = \ln(\text{streamflow}) - \text{center of } \ln(\text{streamflow}) \quad (12)$$

The predefined model corresponding to equation 11 is requested by setting MODNO equal to 1 (record type 7)⁷. A single constituent is to be analyzed (NCONST=1, record type 12)⁸. Phosphorus concentrations specified in the calibration file are in units of milligrams per liter (UCFLAG=1, record type 13); load estimates written to the constituent and individual output files are in units of tons per day (ULFLAG=4, record type 13).

⁷Record types 5 and 6 are omitted when seasonal load calculations are not requested (LDOPT=0, record type 4).

⁸Record type 8 is omitted when the predefined model is not model number 10 or 11; record types 9-11 are omitted when a predefined model is requested.

Header File for Application 1

```

#####
#
# LOADEST Header File
#
# Illinois River at Marseilles, Illinois (Helsel & Hirsch, 2002)
#
#####
1 Illinois River at Marseilles, Illinois (Helsel & Hirsch, 2002)
2 1 | PRTOPT (col.1-5)
3 3 | SEOPT (col.1-5)
4 0 | LDOPT (col. 1-5)
#####
#
# model number, MODNO (col.1-5)
#
#####
7 1
#####
#
# number of constituents, NCONST (col.1-5)
#
#####
12 1
#####
#
# Unit flags and constituent names, for I=1, NCONST
#
#                               Unit Flags
#CNAME                          Conc   Load
#                               |       |
#####
13 phosphorus                    1     4

```

Figure 4. Header file for Application 1.

A partial listing of the calibration file for the Illinois River data is shown in figure 5. There are 96 observations of streamflow and concentration in the calibration data set. These 96 observations are specified by repeatedly using record type 1 (Section 3.2.3, table 14). Each occurrence of record type 1 includes the date and time of the sample (CDATE and CTIME), the observed streamflow (CFLOW), and the phosphorus concentration (CCONC)⁹.

⁹The CADDL field is omitted from record type 1 when a predefined model is selected (MODNO≠99; record type 7, header file).

Calibration File for Application 1

```
#####
#
#
# LOADEST Calibration File
#
# Illinois River at Marseilles, Illinois (Helsel & Hirsch, 2002)
#
# Note: Sample dates (CDATE) were extracted from the decimal times given by
#       Helsel and Hirsch (2002). Sample times (CTIME) are arbitrarily set to '1200'
#       and do not reflect the actual sample times.
#
#####
#
#
# CDATE      CTIME      CFLOW      CCONC
#
#####
#
19741113      1200      6860.      1.40
19741216      1200      6640.      0.69
19750114      1200      9870.      0.74
19750212      1200      8690.      1.00
19750317      1200      11300.     0.55
19750415      1200      9010.      0.50
19750515      1200      15400.     0.57
19750616      1200      9560.      0.55
19750716      1200      7760.      0.59
19750814      1200      4970.      0.63
19750916      1200      8070.      0.37
19751015      1200      2980.      1.30
19751113      1200      2930.      0.92
19751216      1200      12500.     0.61
19760114      1200      3550.      0.28
19760212      1200      9130.      0.54
19760317      1200      33300.     0.95
19760415      1200      10700.     0.31
19760514      1200      11200.     0.37
19760616      1200      10900.     0.64
19760715      1200      5840.      0.33
19760813      1200      3480.      0.47
19760915      1200      2680.      0.68
19761113      1200      2720.      0.79
19761216      1200      3130.      0.91
```

Continued...

Figure 5. Partial listing of the calibration file for Application 1 (calibration data for 1977–1985 not shown).

A partial listing of the estimation file for this application is shown in figure 6. A single observation of the data variables (time and streamflow) is specified for each day (NOBSPD=1, record type 1). Record type 2 is used repeatedly to specify the date (EDATE), time (ETIME), and streamflow (EFLOW) for each observation (fig. 6)¹⁰.

¹⁰The EADDL field is omitted from record type 2 when a predefined model is selected (MODNO≠99; record type 7, header file).

Estimation File for Application 1

```
#####
#
# LOADEST Estimation File
#
# Illinois River at Marseilles, Illinois (Helsel & Hirsch, 2002)
#
# Note: Sample dates (CDATE) were extracted from the decimal times
#       given by Helsel and Hirsch (2002). Sample times (CTIME) are
#       arbitrarily set to '1200' and do not reflect the actual
#       sample times.
#
#####
#
# Number of observations per day, NOBSPD (col. 1-5)
#
#####
1 1
#####
#
# EDATE      ETIME      EFLOW
#
#####
2 19741113    1200      6860.
2 19741216    1200      6640.
2 19750114    1200      9870.
2 19750212    1200      8690.
2 19750317    1200     11300.
2 19750415    1200      9010.
2 19750515    1200     15400.
2 19750616    1200      9560.
2 19750716    1200      7760.
2 19750814    1200      4970.
2 19750916    1200      8070.
2 19751015    1200      2980.
2 19751113    1200      2930.
2 19751216    1200     12500.
2 19760114    1200      3550.
2 19760212    1200      9130.
2 19760317    1200     33300.
2 19760415    1200     10700.
2 19760514    1200     11200.
2 19760616    1200     10900.
2 19760715    1200      5840.
2 19760813    1200      3480.
2 19760915    1200      2680.
2 19761113    1200      2720.
2 19761216    1200      3130.
```

Continued...

Figure 6. Partial listing of the estimation file for Application 1 (estimation data for 1977–1985 not shown).

Calibration results are shown in table 18 and in figure 7. Because the data set is uncensored, model coefficients for the AMLE and MLE methods are identical (table 18), and the same model fit results (fig. 7). A slightly different fit results from the LAD regression (table 18 and fig. 7). Inspection of figure 7 reveals a linear relation between log load and log streamflow, with a fairly random distribution of observations around the regression line. This randomness is further illustrated in figure 8, where model residuals are plotted with load. As shown in the figure, positive and negative residuals are scattered, without any clear pattern or trends. The model residuals therefore appear to be independent and homoscedastic (having constant variance), two assumptions of the model fit. Further investigation of these assumptions may be warranted, however, and additional tests are presented elsewhere (Draper and Smith, 1998; Helsel and Hirsch, 2002).

Table 18. Calibration results, Application 1.

Method	Model coefficients (equ. 11)	
	a_0	a_1
AMLE	9.336	0.761
MLE	9.336	0.761
LAD	9.331	0.775

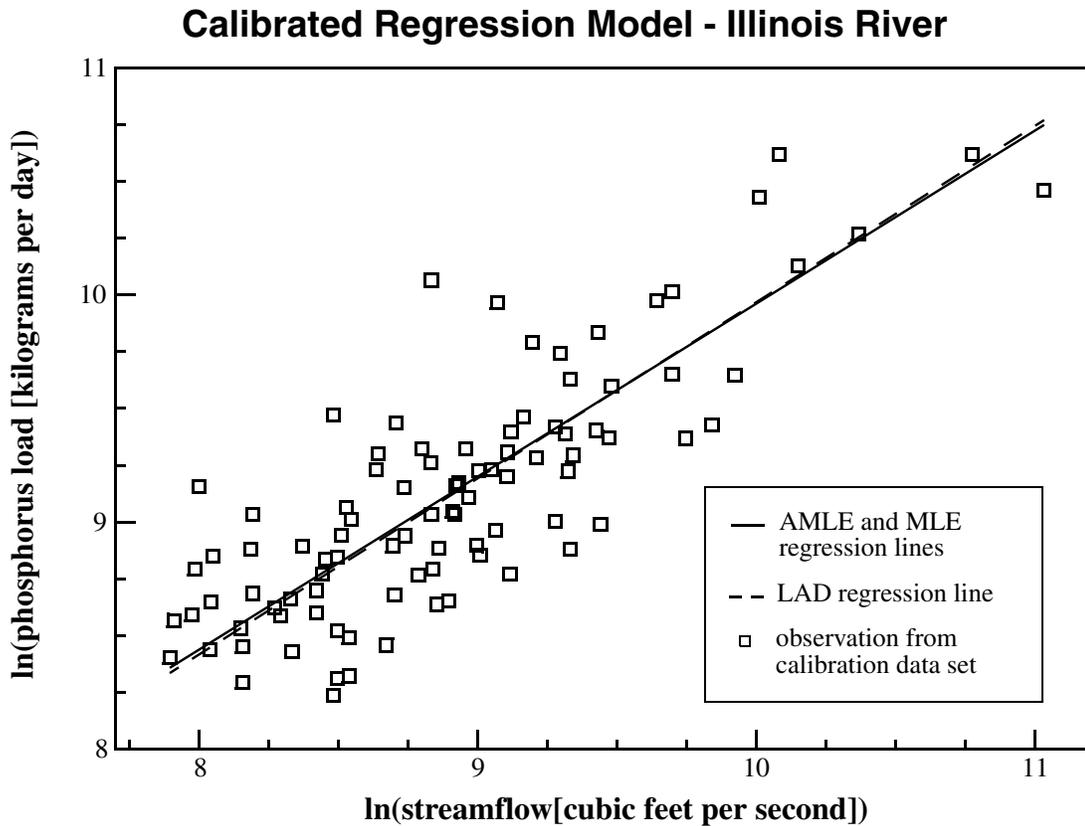


Figure 7. Calibrated regression model for the Illinois River showing log load as a function of log streamflow.

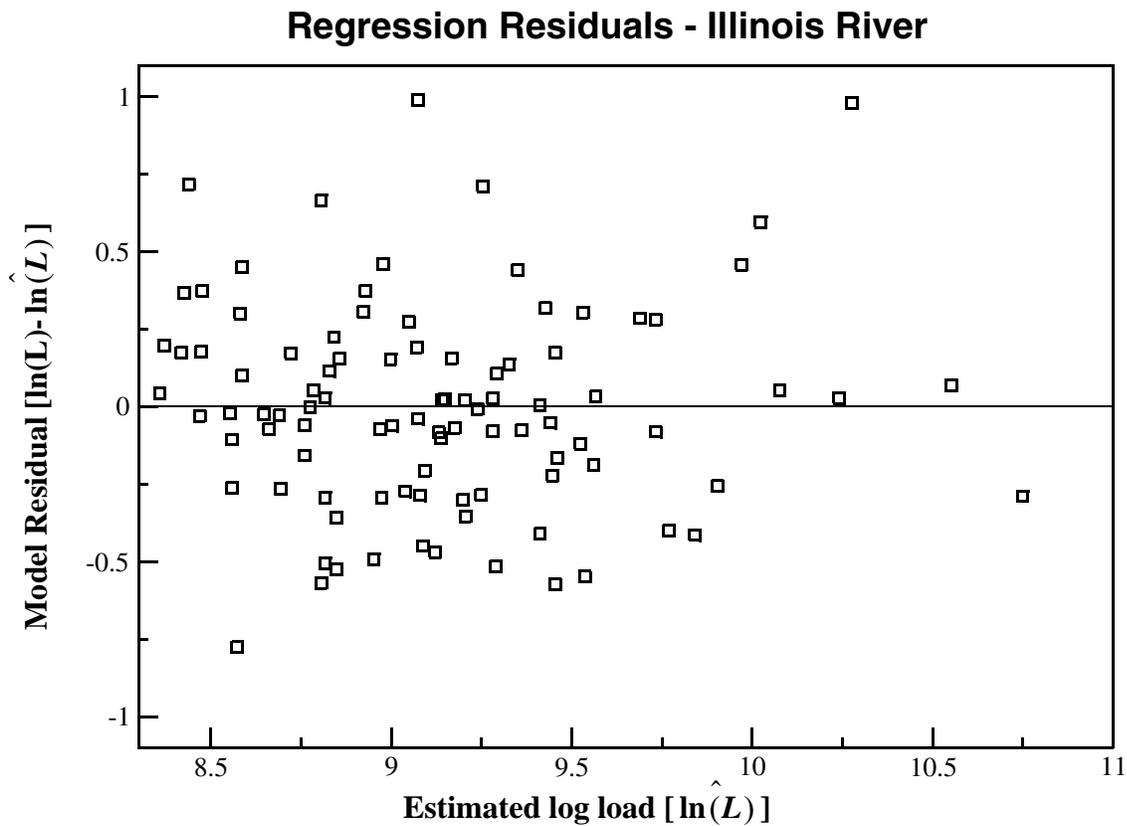


Figure 8. Estimates of log load versus model residuals from AMLE regression, Illinois River (because the data set is uncensored, MLE residuals are identical to those shown here for AMLE).

As discussed in Section 2.2.4, the AMLE and MLE load estimation methods assume a normal distribution of model residuals. As a check on this assumption, a normal probability plot is constructed using the first two columns of the residual output file (fig. 9). The linearity of the plot suggests that the residuals follow a normal distribution. This linearity is supported by a Probability Plot Correlation Coefficient (PPCC) of 0.99 (where a value of 1.00 represents a perfect linear relationship; Helsel and Hirsch, 2002).

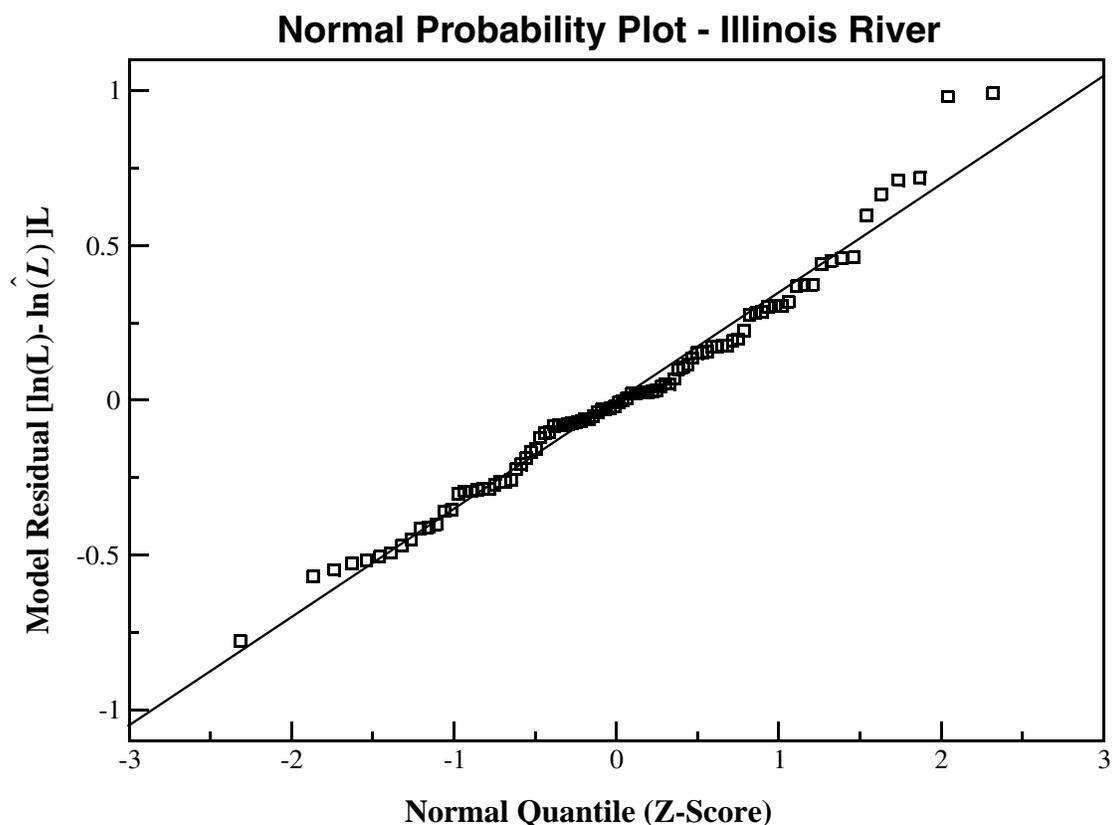


Figure 9. Normal probability plot for the model residuals of the AMLE regression, Illinois River (because the data set is uncensored, MLE residuals are identical to those shown here for AMLE).

Mean load estimates for the entire estimation period are presented in table 19. Because the calibration data set is uncensored, the AMLE method reduces to MLE, and identical load estimates result. The mean LAD load estimate differs from the other two estimates because the model coefficients differ and the Duan bias correction factor is used.

All of the load estimation methods result in load estimates that are within 2 percent of the “true” load (table 19). Of the three, the LAD estimate has the lowest error relative to the true load. Note, however, that the “true” load is rarely known, and other criteria for selecting the appropriate method usually are employed. As discussed in Section 2.2.4, the AMLE load estimate would be used in the case considered here, due to the apparent normality of the residuals (fig. 9).

Table 19. Mean load estimates, Application 1.

Method	Mean load [tons/day]	Percent error
“true” load	12.63	
AMLE	12.38	1.98
MLE	12.38	1.98
LAD	12.49	1.11

4.2 Application 2: Analysis of an Uncensored Constituent using a Seasonal Model

The purpose of this application is to illustrate the use of LOADEST for constituents that exhibit a seasonal loading pattern. In addition, this application goes beyond Application 1 to estimate loads based on streamflow data that are not included in the calibration data set. This extension of the loading record is the primary use of LOADEST. The ability to define seasons for which mean load estimates are calculated also is illustrated in this example.

Many constituents are subject to seasonal loading patterns that are driven by external factors such as the length of the growing season, air temperature, and anthropogenic activities. Some of the predefined models (Section 3.2.2, table 7) implement sine and cosine terms to reflect the seasonality of constituent load (for example, see equ. 14, Section 4.3). These models are applicable to constituents that change continuously over the seasonal cycle, but are not appropriate for other constituents that are subject to an abrupt seasonal change. Such is the case for pesticides, where springtime application results in a distinct change in the loading pattern for a discrete period of time. When such changes are evident in the data, it often is beneficial to use dummy explanatory variables (Judge and others, 1988; Helsel and Hirsch, 2002) to handle the abrupt change. These dummy variables are incorporated in predefined models number 10 and 11 (Section 3.2.2, table 7). In this example, data from the St. Joseph River near Newville, Ind., is used with model number 10 to estimate atrazine loads. Observations of atrazine load show two distinct relations with regard to streamflow; for a given streamflow, atrazine loads are elevated during the 3-month period following pesticide application, and dramatically lower thereafter. This two-tiered loading response is modeled using the dummy explanatory variables present in model 10:

$$\ln(\hat{L}) = a_0 + a_1 per + a_2 \ln Q + a_3 \ln Q per \quad (13)$$

where $\ln Q$ is defined in Application 1 (Section 4.1, equ. 12) and per is a dummy variable set to a binary value (0 or 1). For the case considered here, per is set equal to 1 for the 3-month period (May–July) following atrazine application. The regression line for estimation during this time period will have an intercept equal to $a_0 + a_1$, and a slope of $a_2 + a_3$. For all other time periods (August–April), per will equal 0 and the regression equation corresponds to the simple model used in Application 1 (Section 4.1, equ. 11).

The control, calibration, and estimation input files for the St. Joseph River are in the same format as the corresponding files in Application 1 (Section 4.1, figs. 3, 5, and 6). The header file is shown on figure 10. Specification of record types 1–3 follows that described for Application 1. Record type 4 is used to request the calculation of mean load for a set of user-defined seasons (LDOPT=1). Record types 5 and 6 then are used to define the number of seasons (NSEAS=4) and the begin and end dates for each season (SBEG and SEND, respectively). Because there are multiple seasons, record type 6 is used repeatedly to define the time periods for each season. Each season is defined as a 3 month period (February 1–April 30, May 1–July 31, August 1–October 31, and November 1–January 31), with the begin and end dates specified in MMDD format. The regression model corresponding to equation 13 is requested using record type 7 (MODNO=10). When this model (or model 11) is selected, the time period for the dummy variable [per , equ. 13] is set using record type 8. For the case considered here, loads are elevated from May through July, so PBMON is set to 5 (=May) and PEMON is set to 7 (=July) such that per equals 1 during that period and 0 thereafter. As in Application 1, record types 12 and 13 are used to specify the number of constituents, the constituent name, and the unit flags¹¹.

¹¹Record types 9-11 are omitted when a predefined model is requested.

Header File for Application 2

```

Record Type #####
#
# LOADEST Header File
#
# St.Joseph River near Newville, IN (Station # 04178000)
#
#####
1 St.Joseph River near Newville, IN (Station # 04178000)
2 1 | PRTOPT (col.1-5)
3 3 | SEOPT (col.1-5)
4 1 | LDOPT (col. 1-5)
#####
#
# setup user defined seasons (LDOPT = 1 or 3 only)
#
# number of user-defined seasons (NSEAS) and begin and end
# dates for each season (SBEG,SEND)
#
#####
5 4 | NSEAS (col.1-5)
#
#SBEG SEND (I=1,NSEAS)
#####
6 0201 0430
6 0501 0731
6 0801 1031
6 1101 0131
#####
#
# model number, MODNO (col.1-5)
#
#####
7 10
#####
#
# beginning and ending months for period (MODNO = 10 or 11 only)
#
#
#PBMN PEMON
#
#####
8 5 7
#####
#
# number of constituents, NCONST (col.1-5)
#
#####
12 1
#####
#
# Unit flags and constituent names, for I=1, NCONST
#
#
#CNAME Unit Flags
# Conc Load
# | |
#####
13 atrazine 2 3

```

Figure 10. Header file for Application 2.

As noted in the discussion of equation 13, use of the dummy variable results in two distinct regression lines (fig. 11). This two-tiered loading response results in higher estimates of mean load for the time period following the atrazine application (May–July, the second user-defined season), relative to the other user-defined seasons (fig. 12). Although the atrazine analysis presented in figures 11 and 12 appears promising, two issues remain cause for concern. First, the calibrated model for May–July appears to underestimate the atrazine load when streamflow is elevated (for higher values of streamflow, observed loads plot above the regression line; fig. 11). Second, inspection of the constituent output file (Part IIa) indicates that some of the streamflow values used for estimation exceed the maximum streamflow in the calibration data set. As a result, some of the load estimates require extrapolation beyond the range of streamflow used for calibration. Because of these concerns, additional data collection and model formulation activity may be warranted.

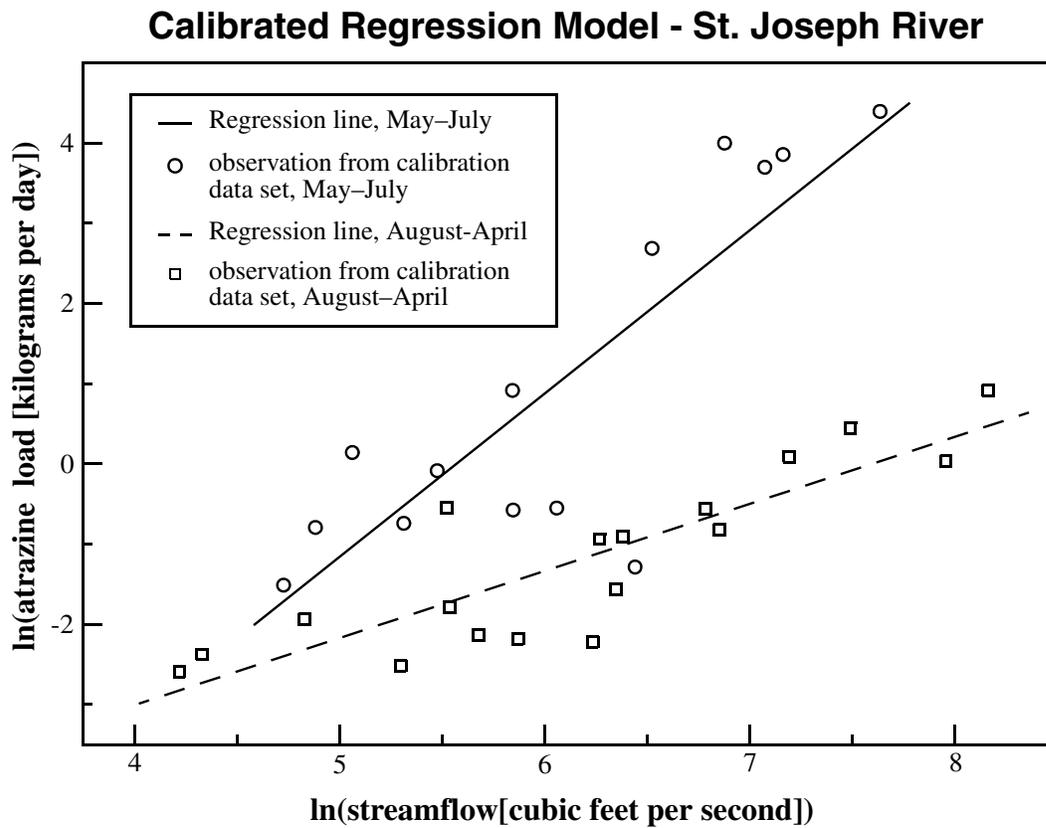


Figure 11. Calibrated regression model for the St. Joseph River showing log load as a function of log streamflow.

Seasonal Load Estimates - St. Joseph River

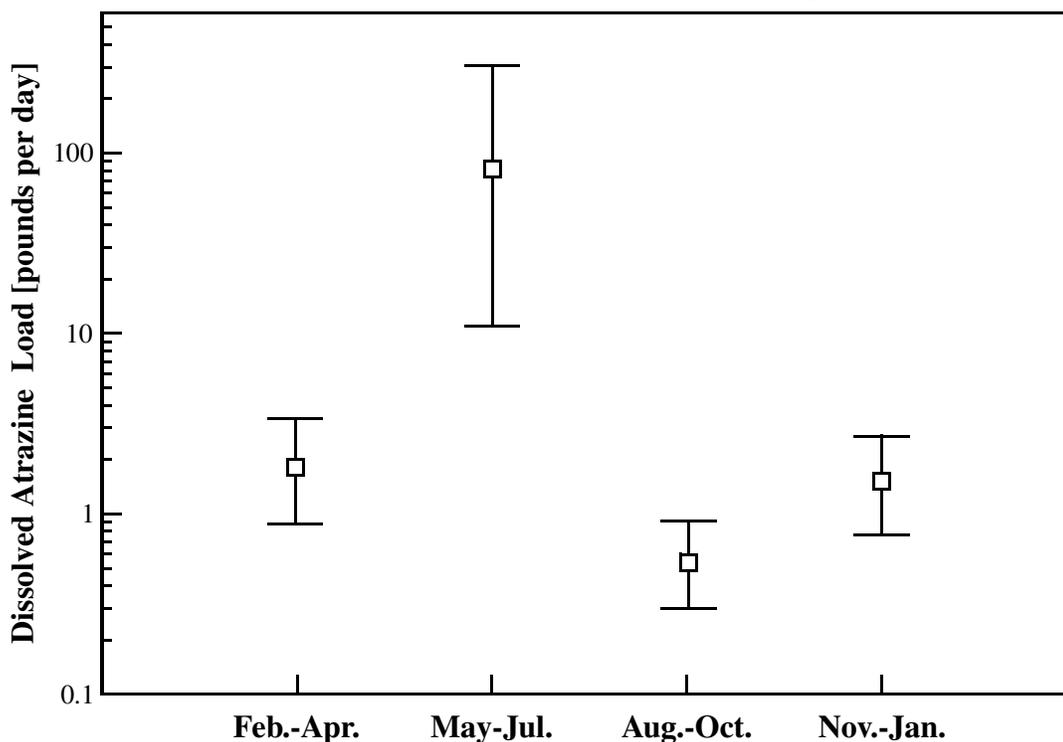


Figure 12. Seasonal AMLE load estimates for the St. Joseph River. Symbols denote mean load for specified season; lines represent 95-percent confidence intervals.

4.3 Application 3: Analysis of a Censored Constituent using a Seasonal Model

This application illustrates the “7-parameter model,” a predefined model that has been shown to perform well for loading analyses of constituents in large (> 100 miles²) watersheds (Cohn and others, 1992a). In addition, the application illustrates the use of LOADEST when portions of the calibration data set are subject to censoring.

As in the previous example, a constituent with a seasonal loading pattern is considered here. In this case, constituent concentrations are assumed to vary in a continuous manner, as opposed to the abrupt changes considered in Application 2. Several of the predefined models (models 4 and 6–9; Section 3.2.2, table 7) use a first-order Fourier series (sine and cosine terms) to consider seasonality. In this application, the 7-parameter model (model 9) is developed for nutrient loading on the Potomac River. The 7-parameter model is given by:

$$\ln(\hat{L}) = a_0 + a_1 \ln Q + a_2 \ln Q^2 + a_3 \sin(2\pi dtime) + a_4 \cos(2\pi dtime) + a_5 dtime + a_6 dtime^2 \quad (14)$$

where $\ln Q$ is defined in Application 1 (Section 4.1, equ. 12) and $dtime$ is equal to decimal time minus the center of decimal time (Section 2.3). Within the model, explanatory variables one and two account for the dependence on streamflow, explanatory variables three and four are a first-order Fourier series to account for seasonal variability, and explanatory variables five and six account for the time trend.

Equation 14 is calibrated using orthophosphate data collected near USGS gaging station 01646580 on the Potomac River. The calibration data set includes 198 observations of flow and concentration collected from 1978 to 1990; 46 of the observations are below the laboratory detection limit, resulting in a censored data set. Control and estimation input files for

this application are analogous to those shown in Applications 1 and 2 and are not shown here. The header file is similar to that of Application 1; one difference of note is record type 4, where monthly load estimates are requested (LDOPT=2). A partial listing of the calibration file is shown in figure 13. As with the previous examples, record type 1 is used repeatedly to specify the date and time of the sample (CDATE and CTIME), the observed streamflow (CFLOW), and the constituent concentration (CCONC)¹². For uncensored observations, CCONC is set equal to the observed concentration (CCONC ≥ laboratory detection limit; Section 3.2.3, table 15). Censored observations are flagged by placing a less-than sign (<) at the start of the CCONC field; the less-than sign is immediately followed by the laboratory detection limit (bold fields, fig. 13).

Calibration File for Application 3

```
#####
#
# LOADEST Calibration File
#
# Potomac River at Little Falls/Chain Bridge (Station # 01646580)
#
# Obtained from Tim Cohn, 15 September 2003
#
#####
#
#CDATE      CTIME      CFLOW      CCONC
#
#####
19781219    1000      5700.0     0.040
19790301    1230      78200.0    0.030
19790319    1230      16600.0    0.020
19790327    1000      48300.0    0.010
19790426    1300      10100.0    <0.010
19790523    1045      9960.0     <0.010
19790605    1030      33900.0    0.010
19790613    1030      20600.0    0.030
19790717    1030      4660.0     <0.010
19790815    1030      5000.0     <0.010
19790906    1330      51200.0    0.030
19790906    2000      51200.0    0.030
19790907    0745      84000.0    0.030
19790907    1100      84000.0    0.040
19790907    1430      84000.0    0.040
19790907    1830      84000.0    0.020
19790909    1415      38900.0    0.040
19791011    1130      69700.0    0.040
19791127    1100      12900.0    <0.010
19791212    1030      8830.0     <0.010
```

Continued...

Figure 13. Partial listing of the calibration file for Application 3 (calibration data for 1980-1990 not shown). Fields in bold indicate observations that are censored (CCONC= <0.010).

AMLE regression results are shown in table 20 (for a discussion of other calibration issues, including residual analyses and selection of AMLE, see Applications 1 and 2). Estimated values and the associated standard deviations are shown for each model coefficient. Coefficient t-ratios are equal to the estimated value divided by its standard deviation; when the absolute value of the t-ratio is less than 1, the estimated coefficient is highly uncertain (coefficients a_2 and a_4 , table 20). Similarly, high values of the likelihood ratio P-value for a_2 and a_4 indicate that the estimated coefficient may not be significantly different than zero. Additional details on t-ratios, P-values, and tests of significance are given in Draper and Smith (1998) and Helsel and Hirsch (2002).

¹²The CADDL field is omitted from record type 1 when a predefined model is selected (MODNO≠99; record type 7, header file).

Table 20. AMLE model coefficients, Application 3.

Coefficient: Explanatory Variable	Centered by:	Value	Standard deviation	t-ratio	P-value
a_0 : constant	not centered	6.8560	0.2376	28.85	2.578E-61
a_1 : $\ln Q$	9.2926	1.0802	0.0959	11.26	4.448E-23
a_2 : $\ln Q^2$	9.2926	-0.0539	0.0648	-0.83	3.893E-01
a_3 : $\sin(2\pi dtime)$	1984.513	0.8545	0.1711	4.99	6.507E-07
a_4 : $\cos(2\pi dtime)$	1984.513	0.1236	0.1499	0.82	4.021E-01
a_5 : $dtime$	1984.513	-0.1191	0.0276	-4.31	1.097E-05
a_6 : $dtime^2$	1984.513	-0.0188	0.0113	-1.66	8.454E-02

Unlike the previous examples, use of the 7-parameter model introduces several explanatory variables and the effects of multicollinearity (Section 2.3) should be considered. An initial assessment of multicollinearity (or the lack thereof) may be conducted by inspecting the cross-correlation matrix for the explanatory variables, shown in table 21. Collinearity between the model's linear and quadratic terms ($\ln Q$ and $\ln Q^2$; $dtime$ and $dtime^2$) has been eliminated by centering log streamflow and decimal time (Section 2.3; see correlation coefficients of 0.00, table 21). Most of the other correlation coefficients also are small, indicating that the associated explanatory variables are not highly correlated. One exception is the relatively high correlation (-0.5698) between model coefficients a_1 ($\ln Q$) and a_3 [$\sin(2\pi dtime)$]. This high correlation coefficient may warrant further diagnostic tests for multicollinearity, and investigation of alternate models (for example, see the variance inflation factor described by Helsel and Hirsch, 2002).

Table 21. Correlation between explanatory variables, AMLE regression, Application 3.

	$a_1:\ln Q$				
$a_2:\ln Q^2$	0.0000	$a_2:\ln Q^2$			
$a_3:\sin(2\pi dtime)$	-0.5698	0.3817	$a_3:\sin(2\pi dtime)$		
$a_4:\cos(2\pi dtime)$	-0.2751	0.0561	0.1990	$a_4:\cos(2\pi dtime)$	
$a_5:dtime$	-0.2082	-0.0795	0.0269	-0.1368	$a_5:dtime$
$a_6:dtime^2$	0.1071	-0.1165	-0.0257	0.1299	0.0000

The 7-parameter model is used with daily values of streamflow to estimate AMLE loads for the 1980 calendar year. The mean load for the entire estimation period (366 load estimates for the 1980 calendar year) is 1,580 kilograms per day, with a standard error of prediction of 319 kilograms per day. Mean monthly load estimates and the associated 95 percent confidence intervals are shown in figure 14.

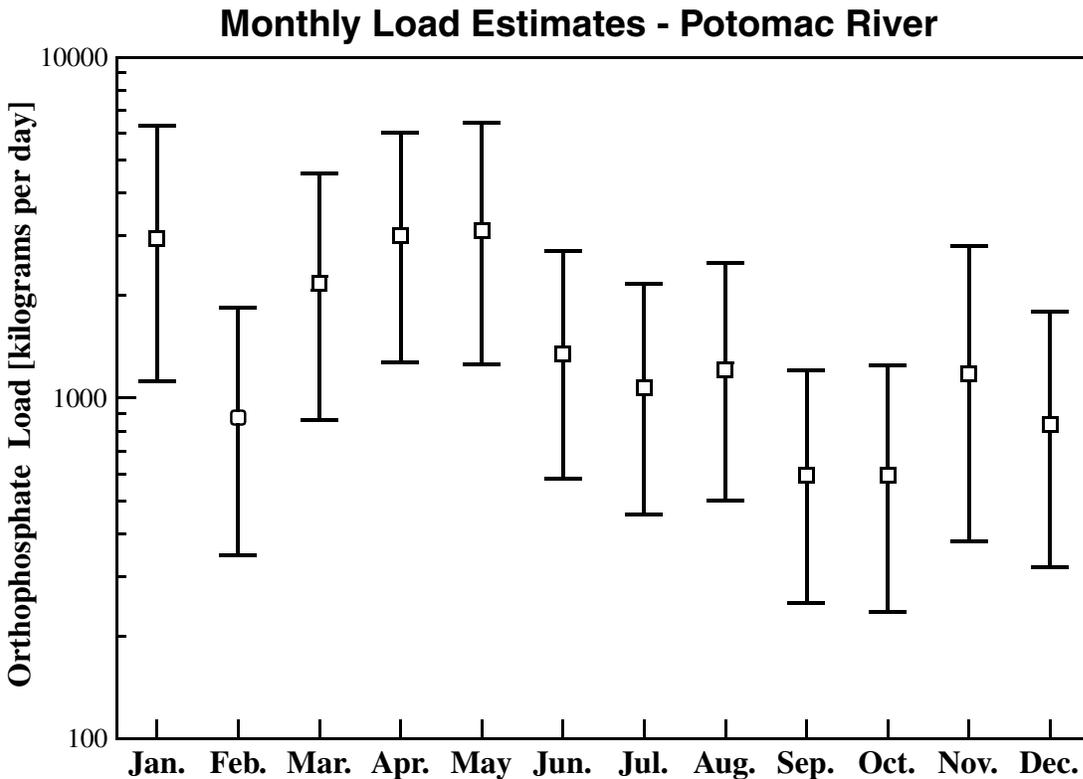


Figure 14. Monthly AMLE load estimates for the Potomac River. Symbols denote mean load for the specified month; lines represent 95-percent confidence intervals.

4.4 Application 4: Multiple Constituents with Automated Model Selection

This application illustrates the automated model selection option (Section 2.4) and the format of the input files when multiple constituents are considered. Four constituents are analyzed to illustrate the use of LOADEST when data sets are uncensored, slightly censored, highly censored, or have missing observations.

The data used in this application were collected at USGS gaging station 03374100, on the White River at Hazleton, Ind. Forty-five water-quality samples collected between October 1992 and September 1994 are used in conjunction with observed streamflow data to calibrate the regression model. The calibrated model is then used to develop load estimates using 730 observations of daily streamflow from the same time period. Constituents considered in this example include butylate, atrazine, alachlor, and suspended sediment.

Sample input files corresponding to the White River analysis generally correspond to those presented in the foregoing applications. Due to the complexity added by the consideration of multiple constituents, the header and calibration files are shown in figures 15 and 16. Figure 15 depicts the header file, where specification of record types 1-3 follows that of the previous applications. Record type 4 is used to request both seasonal and monthly load estimation (LDOPT=3). Record types 5 and 6 are used to define the seasons as in Application 2. The automated model selection option is requested using record type 7 (MODNO=0), and record type 12 is used to set the number of constituents (NCONST=4)¹³. Record type 13 is used once for each of the constituents to specify the constituent name (CNAME), the concentration unit flag (UCFLAG), and the load unit flag (ULFLAG).

¹³Record type 8 is omitted when the predefined model is not model number 10 or 11; record types 9–11 are omitted when a predefined model is requested.

Header File for Application 4

```

#####
#
# LOADEST Header File
#
# White River at Hazleton IN (Station #03374100)
#
#####
1 White River at Hazleton IN (Station #03374100)
2 1 | PRTOPT (col.1-5)
3 3 | SEOPT (col.1-5)
4 3 | LDOPT (col. 1-5)
#####
#
# setup user defined seasons (LDOPT = 1 or 3 only)
#
# number of user-defined seasons (NSEAS) and begin and end
# dates for each season (SBEG,SEND)
#
#####
5 4 | NSEAS (col.1-5)
#
#SBEG SEND (I=1,NSEAS)
#####
6 0301 0531
6 0601 0831
6 0901 1130
6 1201 0229
#####
#
# model number, MODNO (col.1-5)
#
#####
7 0
#####
#
# number of constituents, NCONST (col.1-5)
#
#####
12 4
#####
#
# Unit flags and constituent names, for I=1, NCONST
#
#
# Unit Flags
#CNAME Conc Load
# | |
#####
13 butylate 2 3
13 atrazine 2 3
13 alachlor 2 3
13 suspended sediment 1 4

```

Figure 15. Header file for Application 4.

The calibration file for the White River data is shown in figure 16. There are 45 observations of streamflow and concentration in the calibration data set. These 45 observations are specified by repeatedly using record type 1. Each occurrence of record type 1 includes the date and time of the sample (CDATE and CTIME), the observed streamflow (CFLOW), and the concentration for the four constituents (CCONC repeats horizontally)¹⁴. Specification of CCONC is dependent on the given constituent and observation. For atrazine, all of the concentrations are uncensored (CCONC \geq laboratory detection limit) and the specified value of CCONC equals the observed concentration (Section 3.2.3, table 15). For alachlor, most of the concentrations are uncensored, but two of the observations are below the detection limit. These censored observations are flagged by placing a less-than sign (<) at the start of the CCONC field (bold fields, fig. 16). For butylate, the data set is highly censored, and 23 observations fall below the detection limit (bold fields, fig. 16). The final constituent, suspended sediment, has three missing observations that are flagged by setting CCONC equal to -9999.0 (CCONC \leq 0.0; bold fields, fig. 16); the remaining observations are uncensored.

¹⁴The CADDL field is omitted from record type 1 when a predefined model is selected (MODNO \neq 99; record type 7, header file)

Calibration File for Application 4

```
#####
#
# LOADEST Calibration File
#
# White River at Hazleton IN (Station #03374100)
#
# Obtained from Charlie Crawford, 5 July 2001
#
#####
#
#
# CCONC (J=1,NCONST)
#CDATE      CTIME    CFLOW      buty.      atra.      alach.      spsed.
#####
19921019    1200      4290.0      0.004      0.280      0.016      90.5
19921118    1145      31400.0     0.004      0.340      0.047      161.0
19921221    1130      7770.0      <0.002     0.180      0.026      -9999.0
19930112    1145      39700.0     0.009      0.230      0.047      -9999.0
19930218    1300      9310.0      <0.002     0.100      0.013      -9999.0
19930330    1400      20900.0     <0.002     0.150      0.034      98.0
19930412    1400      31300.0     <0.002     0.540      0.025      118.0
19930426    1300      18700.0     <0.002     0.200      0.018      90.0
19930503    1400      16800.0     0.010      1.100      0.150      135.0
19930510    1500      18200.0     0.009      1.600      0.070      140.0
19930517    1430      15600.0     0.027      1.300      0.093      154.0
19930524    1300      9150.0      0.042      4.700      0.740      90.0
19930601    1300      6330.0      0.017      1.900      0.170      125.0
19930607    1500      10400.0     0.024      9.400      0.540      368.0
19930614    1400      9760.0      0.024      12.800     2.100      254.0
19930621    1400      12600.0     0.006      10.500     1.200      237.0
19930628    1230      9160.0      0.004      7.800      0.670      153.0
19930706    1400      21200.0     0.004      3.400      0.280      344.0
19930712    1430      11100.0     <0.002     2.900      0.280      222.0
19930719    1400      11200.0     0.006      2.100      0.120      269.0
19930726    1430      6090.0      <0.002     1.100      0.055      125.0
19930802    1430      4530.0      <0.002     0.940      0.034      118.0
19930809    1600      3560.0      0.002      0.670      0.027      73.0
19930816    1300      6910.0      0.003      0.590      0.031      274.0
19930823    1330      23800.0     <0.002     0.450      0.037      196.0
19930830    1330      6700.0      <0.002     0.880      0.054      122.0
19930907    1400      22700.0     <0.002     0.260      0.018      497.0
19930920    1400      7620.0      <0.002     0.310      <0.002     123.0
19931012    1400      13200.0     <0.002     0.210      0.019      252.0
19931118    1150      87900.0     0.004      0.098      0.022      95.0
19931123    1200      113000.0    0.004      0.190      0.048      75.0
19931130    1330      30300.0     <0.002     0.170      0.023      63.0
19931215    1215      19400.0     0.003      0.140      0.021      53.0
19940112    1245      11500.0     <0.002     0.095      0.008      10.0
19940215    1230      13600.0     <0.002     0.060      0.011      43.0
19940305    1215      13400.0     <0.002     0.065      0.012      50.0
19940411    1300      25400.0     0.015      0.740      0.058      478.0
19940509    1330      33100.0     0.026      5.000      0.370      112.0
19940523    1330      11700.0     0.015      4.000      0.300      193.0
19940606    1430      5580.0      <0.002     1.500      0.170      110.0
19940620    1430      4380.0      <0.002     3.900      0.085      136.0
19940705    1330      6320.0      <0.002     6.600      0.670      102.0
19940718    1500      4060.0      <0.002     1.800      0.072      20.0
19940827    1130      1930.0      <0.002     0.410      <0.002     53.0
19940912    1230      1870.0      <0.002     0.430      0.009      53.0
```

Figure 16. Calibration file for Application 4. Bold fields indicate observations that are censored (CCONC= <0.002) or missing (CCONC=-9999.0).

Under the automated model selection option, the White River data are used to calibrate each of the predefined models (Section 3.2.2, table 7) using AMLE. The model with the lowest value of the Akaike Information Criterion (AIC, Section 2.4) is then selected for use in load estimation. For the case of butylate, model number 7 is selected based on its AIC value (atrazine, alachlor, and suspended sediment results are not discussed herein). Inspection of the constituent output file (appendix 3) indicates that model 4 also may be worthy of consideration, based on the Schwarz Posterior Probability Criterion (Section 2.4). Butylate residuals as a function of log streamflow for model number 7 are shown in figure 17. Load estimation based on the selected model results in mean and median butylate loads of 0.68 and 0.175 pounds per day, respectively (appendix 3). A histogram generated using the load estimates from the individual load file is shown as figure 18.

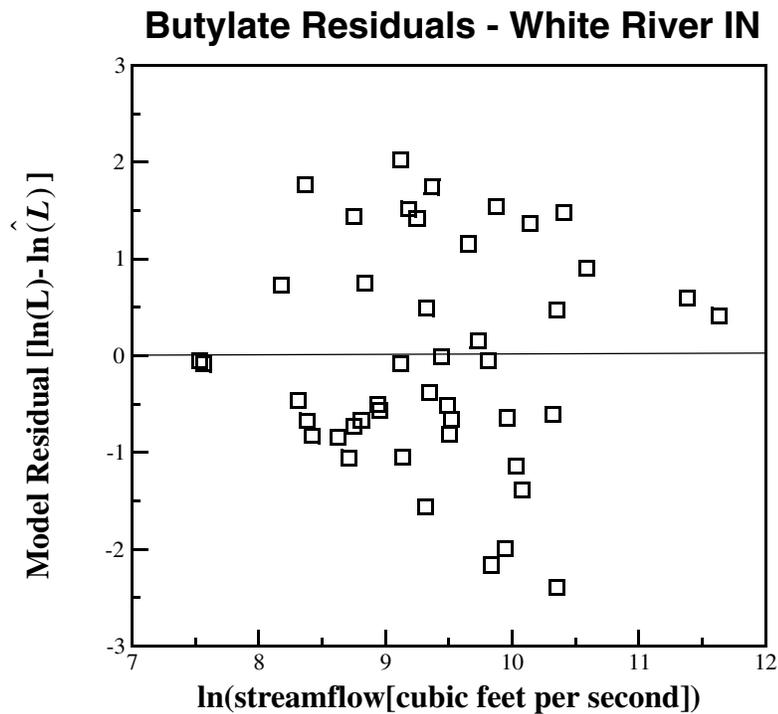


Figure 17. Residual plot for model number 7, White River IN.

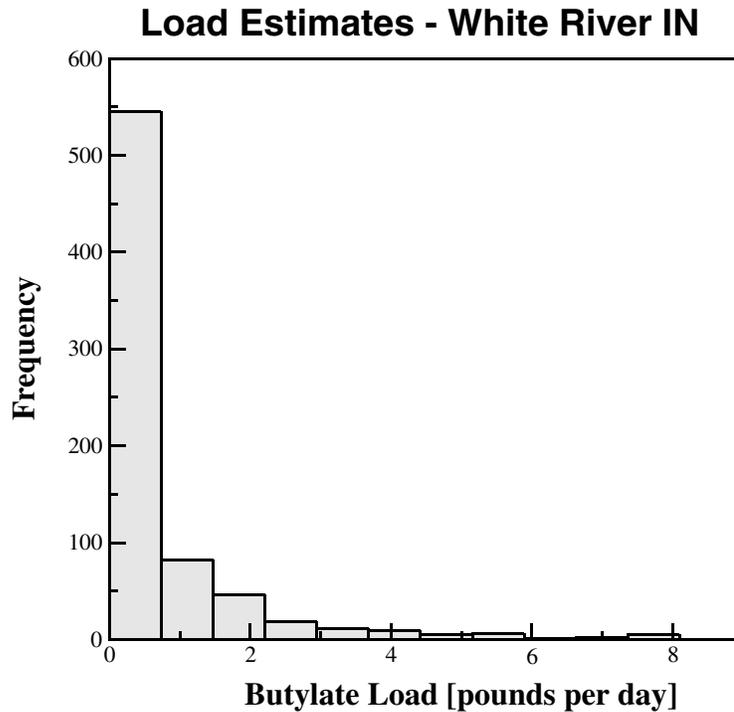


Figure 18. Histogram of estimated AMLE loads, White River IN.

4.5 Application 5: User-Defined Model with an Additional Data Variable

This application illustrates the development of a user-defined model that includes two explanatory variables. The first explanatory variable is based on streamflow, a data variable used extensively in the previous applications. The second explanatory variable is based on specific conductance, an additional data variable that is defined by the user.

Most load estimation efforts to date are based on continuous observations of streamflow and decimal time (Ferguson, 1986; Cohn and others, 1989, 1992a; Crawford, 1996). This reliance on streamflow and decimal time as the only data variables may change in the future because of recent advances in the development of continuous water-quality monitors. These monitors provide observations of surrogate variables (specific conductance, pH, water temperature, dissolved oxygen, and turbidity) that may be useful in the formulation of regression models for load estimation. Following the work of Christensen and others (2000), this example uses 103 observations of streamflow and specific conductance collected from 1995 to 1998 to calibrate a regression model for alkalinity load in the Little Arkansas River, Kansas. The regression model for alkalinity is given by:

$$\ln(\hat{L}) = a_0 + a_1 \ln Q + a_2 \ln(SC) \quad (15)$$

where $\ln Q$ is defined in Application 1 [Section 4.1, equ. 12] and SC is specific conductance in microsiemens per centimeter at 25 degrees Celsius.

Development of equation 15 as a user-defined model is accomplished using several record types in the header file. This is shown in figure 19, where the specification of record types 1–4 generally follows that presented previously. Record type 7 is first used to request a user-defined model (MODNO=99)¹⁵. Record type 9 then is used to define an additional data variable (NADDL=1) so that observations of specific conductance can be used for calibration and estimation¹⁶. The number of explanatory variables is set (NEXPL=2, record type 10) for the two explanatory variables [$\ln Q$ and $\ln(SC)$] in equation 15. Record type 11 completes the specification of the user-defined model by providing the form of each explanatory variable. Field 1 of record type 11 identifies the name of the data variable used to create the explanatory variable (DVNAME) and field 2 specifies the transformation that is applicable to the data variable (TRANS). For the first explanatory variable ($\ln Q$), DVNAME is set equal to Q, the data variable already defined within LOADEST (Section 3.2.2, table 10). For the second explanatory variable [$\ln(SC)$], DVNAME is set to ADDL1, corresponding to the additional data variable defined by record type 9 (ADDL1 is the generic name assigned to specific conductance). Both explanatory variables rely on a log transformation of the data variable (TRANS=LN) in accordance with equation 15. Record types 12 and 13 are used to specify the number of constituents (NCONST), the constituent name (CNAME), and the unit flags.

¹⁵Record types 5 and 6 are omitted when seasonal load calculations are not requested (LDOPT=2, record type 4).

¹⁶Record type 8 is omitted when the model is not model number 10 or 11.

Header File for Application 5

```

#####
#
# LOADEST Header File
#
# Little Arkansas River near Halstead KS (Station # 07143672)
#
# Obtained from Vicki Christensen, 26 September 2003
#
#####
1 Little Arkansas R. nr Halstead KS (Station # 07143672), Alkalinity
2 1      | PRTOPT (col.1-5)
3 3      | SEOPT (col.1-5)
4 2      | LDOPT (col. 1-5)
#####
#
# model number, MODNO (col.1-5)
#
#
7 99
#####
#
# setup user defined model (MODNO = 99 only)
#
# number of additional data variables (NADDL),
# number of explanatory variables (NEXPL) and definitions
# for explanatory variables (DVNAME,TRANS)
#
#####
9 1      | NADDL (col.1-5)
10 2     | NEXPL (col.1-5)
#
#DVNAME  TRANS (I=1,NEXPL)
#####
11 Q      LN
11 ADDL1  LN
#####
#
# number of constituents, NCONST (col.1-5)
#
#####
12 1
#####
#
# Unit flags and constituent names, I=1, NCONST
#
#
#CNAME                                     Unit Flags
#                                           Conc  Load
#                                           |    |
#####
13 alkalinity                             1    3

```

Figure 19. Header file for Application 5.

Use of specific conductance as an additional data variable affects the format of the calibration and estimation files as shown in figures 20 and 21. As with the previous applications, the observations used to develop the regression model are specified using record type 1 of the calibration file (fig. 20). Each occurrence of record type 1 includes the date and time of the sample (CDATE and CTIME), the observed streamflow (CFLOW), the observed specific conductance (CADDL), and the alkalinity concentration (CCONC). The estimation file (fig. 21) is formatted in a similar manner. A single observation of the data variables (time, streamflow, and specific conductance) is specified for each day (NOBSPD=1, record type 1). Record type 2 is used repeatedly to specify the date (EDATE), time (ETIME), streamflow (EFLOW), and specific conductance (EADDL) for each observation (fig. 21).

Calibration File for Application 5

```
#####
#
# LOADEST Calibration File
#
# Little Arkansas River near Halstead KS (Station # 07143672)
#
#####
#
#CDATE      CTIME    CFLOW      CADDL      CCONC
#
#           (S.C.)
#####
19950228    1231      10.0       1425.      248.0
19950324    1301      11.3       1010.      205.0
19950328    0801      190.0      519.       78.0
19950412    0931      13.1       784.       204.0
19950424    1301      22.4       1750.      231.0
19950508    1301     2700.0     98.9       27.0
19950516    1146     123.0     884.       131.0
19950517    1216     651.0     413.       63.0
19950530    1301     9220.0    132.       42.0
19950613    1206     6030.0    208.       58.0
19950621    1116     142.0     910.       170.0
19950705    1111      57.7     1250.      238.0
19950717    1341     35.0     1280.      260.0
19950803    0916     37.4     1200.      242.0
19950905    1001     14.3     828.       248.0
19950919    1031     218.0     724.       118.4
19951017    1111     12.0     1010.      251.8
19951030    1106     12.0     1090.      271.4
19951129    1246     14.1     1440.      302.0
```

Continued...

Figure 20. Partial listing of the calibration file for Application 5 (calibration data for 1996–1998 not shown).

Estimation File for Application 5

```

#####
#
# LOADEST Estimation File
#
# Little Arkansas River near Halstead KS (Station # 07143672)
#
#####
#
# Number of observations per day, NOBSPD (col. 1-5)
#
#####
1 1
#####
#
# EDATE      ETIME      EFLOW      EADDL
#            (S.C.)
#####
2 19990101    1200      44.        1620.
2 19990102    1200      40.        1660.
2 19990103    1200      41.        1620.
2 19990104    1200      38.        1650.
2 19990105    1200      41.        1660.
2 19990106    1200      38.        1630.
2 19990107    1200      35.        1560.
2 19990108    1200      37.        1560.
2 19990109    1200      38.        1600.

```

Continued...

Figure 21. Partial listing of the estimation file for Application 5 (estimation data for 10 January – 31 December not shown).

Calibration results for alkalinity load based on equation 15 generally are favorable. Model coefficients appear to be significant, and the model residuals appear to be normally distributed (for general comments on these issues, see the previous applications; for specific tests of significance and normality, see Helsel and Hirsch, 2002). The two explanatory variables (log streamflow and log specific conductivity) appear to be highly correlated, however, with a cross-correlation coefficient of -0.82 . This high value of the correlation coefficient indicates the presence of multicollinearity, a common pitfall associated with the inclusion of additional data variables. This is especially true of specific conductance and other surrogate variables that may be correlated with streamflow due to the effects of dilution or runoff-enhanced loading. As noted by Helsel and Hirsch (2002), multicollinearity may confound the interpretation of the model coefficients and tests of their significance.

Following calibration, daily values of streamflow and specific conductivity from 1999 are used to estimate loads. Because the 1999 data are not part of the calibration data set, comparison between observed and estimated loads provides a test of the methodology (Christensen and others, 2000). Daily loads for the 1999 calendar year taken from the individual output file are plotted on figure 22. The general agreement between observed and estimated alkalinity load lends confidence to the use of equation 15 for load estimation.

Daily Load Estimates - Little Arkansas River

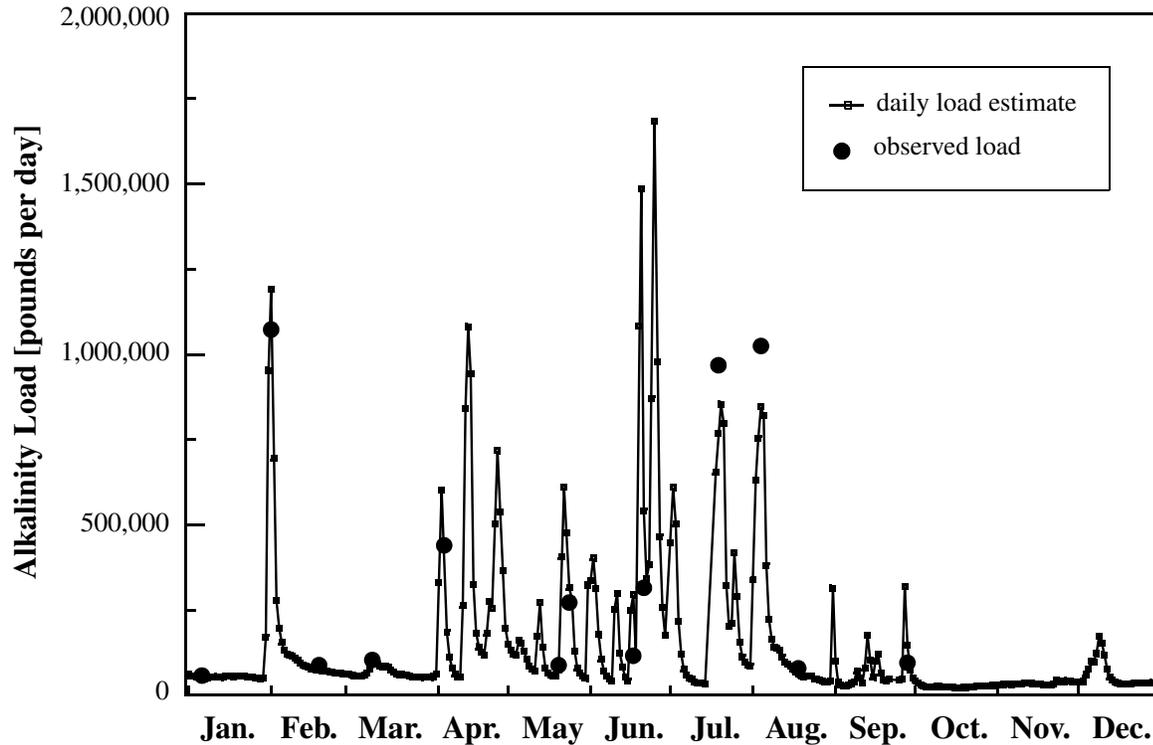


Figure 22. Daily AMLE load estimates for the Little Arkansas River (after Christensen and others, 2000).

4.6 Application 6: Regression Model for Concentration

This application illustrates the use of LOADEST when the primary quantity of interest is constituent concentration, rather than load. As noted in Section 3.4.2, Part Ib of the constituent output file includes a concentration regression that supplements the load regression presented in Part Ia. Although the contents of Part Ib may be useful for users with an interest in constituent concentration, the remaining parts of the output (Part II of the constituent output file, the residuals output file, and the individual load file) are in terms of constituent load. Fortunately, users may “trick” the software into providing additional information related to constituent concentration as shown here. This example pertains to the development of concentration regressions for real-time management decisions (Christensen and others, 2000); another potential application is the use of concentration model residuals to calculate flow-adjusted concentrations (Hirsch and others, 1982).

Data from the Little Arkansas River (Christensen and others, 2000) were used in Application 5 to develop a regression model for constituent load. A similar model for constituent concentration is given by:

$$\ln(\hat{C}) = a_0 + a_1 \ln(\text{streamflow}) + a_2 \ln(SC) \quad (16)$$

where $\ln(\hat{C})$ is an estimate of the log alkalinity concentration. Data from Application 5 are used here to calibrate equation 16. Several changes are made to the input files from Application 5 so that LOADEST provides estimates of constituent concentration rather than load. This is accomplished by setting “streamflow” equal to “1.00” such that:

$$\ln(\hat{L}) = \ln(\hat{QC}) = \ln(\hat{C}) \quad (17)$$

where equation 17 is the left-hand side of equation 4 (Section 2.1). After setting streamflow to 1.0, all LOADEST output labeled as “Load” actually refers to concentration. Estimates of concentration with the corresponding 95 percent confidence intervals thus are readily available from LOADEST output.

The header file for Application 6 is shown in figure 23, where the specification of record types 1–8 follows that from Application 5 (Section 4.5). As shown below, streamflow values in the calibration and estimation files are set so that the left-hand side of equation 4 reduces to $\ln(\hat{C})$ (equ. 17). This presents a small problem in that observations of streamflow are needed to develop the right-hand side of equation 4 [specifically, the $\ln(\text{streamflow})$ term on the right-hand side of equ. 16]. This problem is solved by treating observed streamflow as an additional data variable as was done for specific conductance in Application 5. Record type 9 therefore specifies two additional data variables (NADDL=2 for streamflow and specific conductance), rather than the single additional variable in Application 5. The number of explanatory variables is set (NEXPL=2, record type 10) for the two explanatory variables [$\ln(\text{streamflow})$ and $\ln(SC)$] in equation 16, as in Application 5. Record type 11 completes the specification of the user-defined model by providing the form of each explanatory variable. This differs slightly from Application 5 in that the first explanatory variable [$\ln(\text{streamflow})$] now is based on an additional data variable, ADDL1, rather than Q, the data variable already defined within LOADEST. Field 1 of record type 11 identifies the name of the data variable used to create the explanatory variable (DVNAME), and field 2 specifies the transformation that is applicable to the data variable (TRANS). For the first explanatory variable [$\ln(\text{streamflow})$], DVNAME is set equal to ADDL1, the generic name assigned to the first additional data variable (streamflow). For the second explanatory variable [$\ln(SC)$], DVNAME is set to ADDL2, the generic name assigned to the second additional data variable (specific conductance). Both explanatory variables rely on a log transformation of the data variable (TRANS=LN) in accordance with equation 16. Record types 12 and 13 are used to specify the number of constituents (NCONST), the constituent name (CNAME), and the unit flags. The load unit flag (ULFLAG, Section 3.2.2) in record type 13 is set to 1 so that the calculated “loads” (concentrations) are not converted prior to output¹⁷.

¹⁷As noted in Section 3.4.2, the load regression is in terms of kilograms per day; setting ULFLAG equal to 1 (kilograms per day) results in no conversion of the calculated loads (concentrations).

Calibration File for Application 6

```
#####
#
# LOADEST Calibration File
#
# Little Arkansas River near Halstead KS (Station # 07143672)
#
#####
#
# CDATE      CTIME      CFLOW      CADDL1      CADDL2      CCONC
#           (Flow)      (S.C.)
#####
19950228    1231      0.408735    10.0      1425.      248.
19950324    1301      0.408735    11.3      1010.      205.
19950328    0801      0.408735    190.0     519.       78.
19950412    0931      0.408735    13.1      784.       204.
```

Continued...

Figure 24. Partial listing of the calibration file for Application 6 (observations 5-103 not shown).

Estimation File for Application 6

Record Type	#####					
		# LOADEST Estimation File				
		# Little Arkansas River near Halstead KS (Station # 07143672)				
		#				
		# Number of observations per day, NOBSPD (col. 1-5)				
		#				
	1	1				
		#				
		# EDATE	ETIME	EFLOW	EADDL1	EADDL2
		#		(Flow)	(S.C.)	
	#####					
2	19990101	1200	0.408735	44.	1620.	
2	19990102	1200	0.408735	40.	1660.	
2	19990103	1200	0.408735	41.	1620.	
2	19990104	1200	0.408735	38.	1650.	

Continued...

Figure 25. Partial listing of the estimation file for Application 6 (estimation data for 5 January – 31 December not shown).

5 SOFTWARE GUIDE

This section provides information on how to obtain and install LOADEST (Sections 5.1-5.4). Additional aspects of the software are described in Section 5.5.

5.1 Supported Platforms

Executable binary files are available for both personal computers and workstations running Unix and Windows operating systems¹⁸. A summary of supported operating systems and hardware platforms is provided in table 22. Users with other computer systems will need to compile the source code as described in Section 5.4.

Table 22. Supported systems.

Hardware	Operating system	ARC
Personal computer (Intel x86)	Linux	LIN
Personal computer (Intel x86)	Windows	WIN
Unix workstation (SUN)	Solaris	SOL

Column 3 of table 22 contains ARC, the notation used throughout the remainder of this section to generically denote system architecture. Personal computer users running Linux, for example, should replace the letters “ARC” with “LIN” when following the specific instructions given in Sections 5.2–5.4.

5.2 Software Distribution

LOADEST may be obtained over the Internet using the file transfer protocol (ftp). Potential software users may download source code, hardware-specific executables, and sample input files. Most users will want to download a binary executable and the set of sample input files. These files are stored at the ftp site as zipped tar files (Unix) and self-extracting zip files (Windows). A summary of the downloadable files is presented in table 23.

¹⁸The term “Unix” is used loosely herein to refer to both Linux and Unix-based operating systems. “Windows” as used herein refers to Microsoft Windows 98, Windows NT, Windows XP, Windows 2000, etcetera.

Table 23. Files to download for Unix and Windows systems.

File contents	UNIX (zipped tar file)	Windows (Self-extracting zip file)
LOADEST executable	loadest.ARC.tar.gz	loadWIN.exe
LOADEST source code	loadest.source.tar.gz	loadsrc.exe
Sample input files	loadest.examples.tar.gz	loadexam.exe
Installation batch file	not applicable	install.bat

The files shown in table 23 may be accessed using standard file transfer protocol. The ftp procedure is as follows:

- (1) Log into the ftp depot as anonymous:
 - ftp ftpdcolka.cr.usgs.gov (ftp to depot)
 - anonymous (enter anonymous at the login prompt)
 - anypassword* (enter any password)
- (2) Prepare to download the files:
 - cd pub/loadest (get into appropriate directory)
 - hash (request hash marks)
 - type binary (request binary file transfer)
- (3) Download the sample input files:
 - get loadest.examples.tar.gz (UNIX only)
 - get loadexam.exe (Windows only)
- (4) Download a binary executable, if appropriate¹⁹:
 - get loadest.ARC.tar.gz (UNIX only)
 - get loadWIN.exe (Windows only)
- (5) Download the source code, if appropriate:
 - get loadest.source.tar.gz (UNIX only)
 - get loadsrc.exe (Windows only)
- (6) Download the installation batch file, if appropriate:
 - get install.bat (Windows only)

5.3 Installation

The files described in Section 5.2 may be used to install LOADEST as described here. The installation procedure is designed to create the directory structure shown in figure 26. To begin the installation process, move the downloaded files (Section 5.2) into the *base-directory*. The *base-directory* is a user-selected directory under which the LOADEST directory tree will be placed (fig. 26). After moving the files, follow the instructions for Unix or Windows systems given below.

¹⁹See table 22 and associated text for a definition of "ARC".

LOADEST Directory Structure

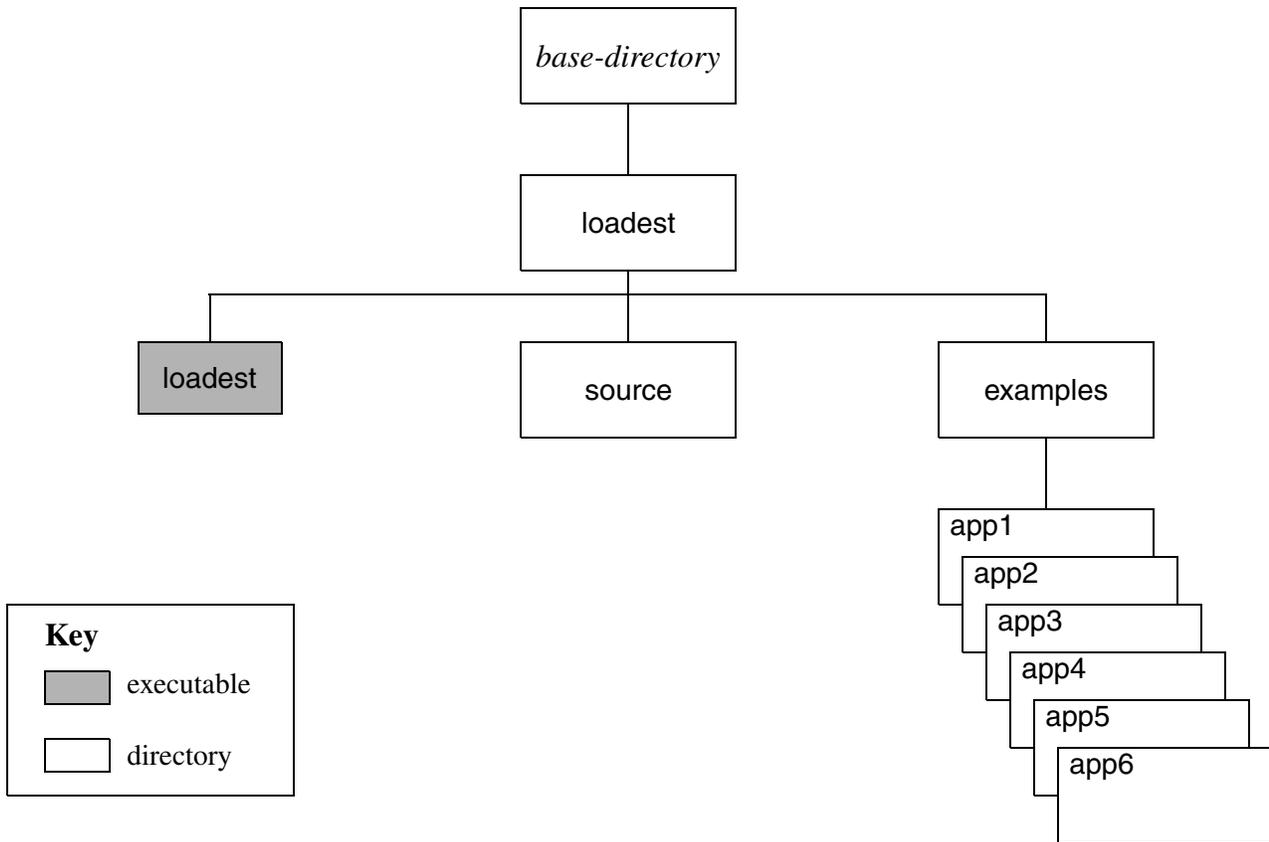


Figure 26: LOADEST directory structure.

Unix Installation Instructions

Installation on Unix systems consists of unzipping the tar files and extracting the directory structure. This is done by issuing the following commands:

- (1) Unzip the tar files:
 `gunzip loadest.examples.tar.gz`
 `gunzip loadest.source.tar.gz` (if appropriate)
 `gunzip loadest.ARC.tar.gz` (if appropriate)
- (2) Extract the directory structure:
 `tar -xovf loadest.examples.tar`
 `tar -xovf loadest.source.tar` (if appropriate)
 `tar -xovf loadest.ARC.tar` (if appropriate)

Windows Installation Instructions

Installation on Windows systems is accomplished by issuing the following command at the MS-DOS prompt²⁰:
install

²⁰Also known as the "Command Prompt" in Windows 2000 (see Programs/Accessories Menu).

Update the user's path (optional)

After completing the installation procedure, the `loadest` executable will reside in *base-directory/loadest/*. Under this arrangement, all LOADEST runs must be initiated from *base-directory/loadest/*. To initiate LOADEST runs from any directory, *base-directory/loadest/* may be added to the user's path.

5.4 Compilation

As discussed in Section 5.1, executable binary files for LOADEST are available for a number of hardware platforms and operating systems. As a result, compilation of the source code may not be required. Compilation is required in the following situations:

- Executable binaries are not available for your specific hardware platform/operating system.
- The dimensions of the analyzed problem exceed the maximum dimensions specified in the include files (Section 5.5.2).
- Modifications have been made to the underlying computer code.

The compilation procedure for Windows systems is not discussed here because it is dependent on the type of compiler used. Under UNIX, the software is compiled using the `make` utility, as described below.

Unix Compilation Instructions

The makefiles used to compile LOADEST are configured to invoke the `f77` Fortran-77 compiler with no command line options. Users requiring other compilers and command line options will need override the default configuration as described in *base-directory/loadest/source/makefile* and *base-directory/loadest/source/makefile.SOL*. LOADEST may be compiled as follows:

- (1) Modify the include files (optional). Changes to the maximum dimensions or logical device assignments may be made by editing the include files described in Section 5.5.2.
- (2) From the *base-directory/loadest/source* directory, enter:
 `make -f makefile.SOL`
if compiling using SPARCompiler Fortran under the SUN Solaris operating system, or
 `make`
if compiling using other compilers and/or operating systems.

When compilation is complete, the `loadest` executable will reside in *base-directory/loadest/*.

5.5 Software Overview

5.5.1 Software Development

LOADEST is written in ANSI standard Fortran-77. The software has been compiled and tested on a variety of hardware platforms and operating systems (Section 5.1). Specific development environments used to compile the source code are presented in table 24.

Table 24. Development environments.

Processor	Operating system	Compiler	Compiler options	ARC
Pentium	Red Hat Linux 7.3	Portland Group Fortran 5.2-1	-fast -tp px -Bstatic -Mdcchk	LIN
Pentium	Windows XP	Portland Group Fortran 5.2-1	-fast -tp px -Mdcchk	WIN
SUN Ultra 10	Solaris 2.6	SPARCompiler Fortran 4.2	-fast -O3 -u	SOL

5.5.2 Include Files

The source code used to develop LOADEST consists of many small subroutines. To facilitate program modification, two include files are used. Use of include files allows program information to be shared between subroutines. This information may be modified by editing the include files rather than each individual routine.

Maximum Dimensions - fmodules.inc

Under Fortran-77, the dimension of each vector and array must be fixed prior to program execution. This requires some knowledge of the maximum dimension for each parameter. Selection of an appropriate size for each parameter is an important task because excessively small values limit program applicability, and excessively large values waste program memory.

To address this problem, the maximum dimensions for the entire software package are defined using an include file, `fmodules.inc`. Increases or decreases in the maximum dimensions are made by editing the include file and compiling the software as described in Section 5.4.

Default values for the maximum dimensions are given in table 25. Some of these dimensions correspond to a user-supplied input variables; this correspondence is shown parenthetically in the third column of the table.

When running the software, the input variables may not exceed the maximum values. The number of constituents (Section 3.2.2, record type 7), for example, may not exceed the maximum value given by `MAXCONST`. When an input value exceeds its given maximum, program execution is terminated and an error message is issued. At this point the user must increase the appropriate maximum value (by editing `fmodules.inc`) and recompile the program.

Table 25. Maximum dimensions and default values from `fmodules.inc`.

Dimension	Default maximum	Maximum number of
MAXADDL	8	Additional data variables (NADDL)
MAXCONST	20	Constituents (NCONST)
MAXOBSC	2,400	Observations for calibration
MAXOBSE	176,000	Observations in the estimation interval
MAXPARMS	7	Model parameters including intercept
MAXSEAS	400	User-defined seasons (NSEAS)

Logical Devices - Ida.inc

In Fortran-77, a unit number is assigned to each file used for input and(or) output. These unit numbers, also known as logical device assignments (*ldas*), must be specified for each read and write operation. Program variables used to store the unit numbers are shared between the input and output subroutines using a Fortran common block. This common block is defined in the include file *lda.inc*.

5.5.3 Fatal Errors and Warnings

The software's input subroutines perform several tests to validate the input data. If fatal errors are detected, an error message is written to *echo.out*, and program execution is terminated. In addition, nonfatal warnings are issued when potential problems with the load estimation procedure are detected. The error-checking capabilities of the software are summarized below.

Fatal Errors

- The estimated values print option, *PRTOPT*, must equal 0 or 1.
- The standard error option, *SEOPT*, must equal 1, 2, or 3.
- The model number, *MODNO*, must be between 0 and 11 (inclusive) or equal to 99.
- The beginning month of the period, *PBMON*, must be between 1 and 12, inclusive.
- The ending month of the period, *PEMON*, must be between 1 and 12, inclusive.
- The ending month of the period, *PEMON*, must be greater than or equal to the beginning month, *PBMON*.
- The concentration unit flag, *UCFLAG*, must equal 1 or 2.
- The load unit flag, *ULFLAG*, must equal 1, 2, 3, or 4.
- The load option, *LDOPT*, must equal 0, 1, 2, or 3.
- The day specified within *CDATE* and *EDATE* must be between 1 and 31, inclusive.
- The month specified within *CDATE* and *EDATE* must be between 1 and 12, inclusive.
- The year specified within *CDATE* and *EDATE* must be later than 1582.
- The time on a given date, *ETIME*, must be in ascending order.
- The constituent concentration, *CCONC*, cannot be both "missing" and "censored".
- The number of observations of the data variables per day, *NOBSPD*, must be 24 or less.
- One or more explanatory variables must be present in a user-defined model (*NEXPL*>0).
- The number of additional data variables, *NADDL*, must not exceed the maximum, *MAXADDL*.
- The number of constituents, *NCONST*, must not exceed the maximum, *MAXCONST*.
- The number of observations in the calibration file must not exceed the maximum, *MAXOBSC*.
- The number of observations in the estimation file must not exceed the maximum, *MAXOBSE*.
- The number of explanatory variables, *NEXPL*, must not exceed the maximum, *MAXPARAMS-1*.
- The number of user-defined seasons, *NSEAS*, must not exceed the maximum, *MAXSEAS*.
- The data variable name, *DVNAME*, must equal *Q*, *DTIME*, or *ADDL*.
- The transformation code, *TRANS*, must equal *NONE*, *SQ*, *SQRT*, *LN*, *LNSQ*, *SIN2P*, *SIN4P*, *SIN6P*, *COS2P*, *COS4P*, or *COS6P*.
- Each day in the estimation interval must have *NOBSPD* observations per day.
- Streamflow values specified in the calibration and estimation files must be greater than zero.

Warnings

Warnings are written to the output files if:

- the maximum estimated concentration exceeds twice the maximum calibration concentration.
- the maximum estimated concentration exceeds 50,000.
- the maximum estimation data set streamflow exceeds the maximum calibration data set streamflow (Load estimates require extrapolation).
- the AIC and SPPC model selection criteria select different models.
- sine or cosine transformations are applied to terms other than decimal time.

REFERENCES

- Bradu, D., and Mundlak, Y., 1970, Estimation in lognormal linear models: *Journal of the American Statistical Association*, v. 65, no. 329, p. 198–211.
- Buchinsky, M., 1994, Changes in the U.S. wage structure 1963–1987—Applications of quantile regression: *Econometrica*, v. 62, no. 2, p. 405–458.
- Christensen, V.G., Jian, X., and Ziegler, A.C., 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. Geological Survey Water-Resources Investigations Report 00-4126, 36 p.
- Cohn, T.A., 1988, Adjusted maximum likelihood estimation of the moments of lognormal populations from type I censored samples: U.S. Geological Survey Open-File Report 88-350, 34 p.
- Cohn, T.A., Delong, L.L., Gilroy, E.J., Hirsch, R.M., and Wells, D.K., 1989, Estimating constituent loads: *Water Resources Research*, v. 25, no. 5, p. 937-942.
- Cohn, T.A., Caulder, D.L., Gilroy, E.J., Zynjuk, L.D. and Summers, R.M., 1992a, The validity of a simple statistical model for estimating fluvial constituent loads—An empirical study involving nutrient loads entering Chesapeake Bay: *Water Resources Research*, v. 28, no. 9, p. 2353–2363.
- Cohn, T.A., Gilroy, E.J., and Baier, W.G., 1992b, Estimating fluvial transport of trace constituents using a regression model with data subject to censoring: *Proceedings of the Joint Statistical Meeting, Boston, August 9–13, 1992*, p. 142–151.
- Cohn, T.A., 1995, Recent advances in statistical methods for the estimation of sediment and nutrient transport in rivers: *Reviews in Geophysics*, v. 33 (Supplement), p. 1117–1124.
- Crawford, C.G., 1991, Estimation of suspended-sediment rating curves and mean suspended-sediment loads: *Journal of Hydrology*, v. 129, p. 331–348.
- Crawford, C.G., 1996, Estimating mean constituent loads in rivers by the rating-curve and flow-duration, rating-curve methods: Bloomington, Ind., Indiana University, Ph.D. dissertation, 245 p.
- Draper, N.R., and Smith, H., 1998, *Applied regression analysis* (3d ed.): New York, John Wiley and Sons, 709 p.
- Duan, N., 1983, Smearing estimate—A nonparametric retransformation method: *Journal of the American Statistical Association*, v. 78, p. 605–610.
- Efron, B., 1982, *The Jackknife, the Bootstrap, and other resampling plans* (Cbms-Nsf Regional Conference Series in Applied Mathematics; 38): Philadelphia, Society for Industrial and Applied Mathematics, 92 p.
- Ferguson, R.I., 1986, River loads underestimated by rating curves: *Water Resources Research*, v. 22, no. 1, p. 74–76.
- Finney, D.J., 1941, On the distribution of a variate whose logarithm is normally distributed: *Supplement to the Journal of the Royal Statistical Society*, v. 7, p. 155–161.
- Gilbert, R.O., 1987, *Statistical methods for environmental pollution monitoring*: New York, Van Nostrand Reinhold, 320 p.
- Gilroy, E.J., Hirsch, R.M., and Cohn, T.A., 1990, Mean square error of regression-based constituent transport estimates: *Water Resources Research*, v. 26, no. 9, p. 2069–2077.
- Harrison, D., 1989, Xgraph Version 11.3.2: Berkeley, University of California.
- Helsel, D.R., and Cohn, T.A., 1988, Estimation of descriptive statistics for multiply censored water quality data: *Water Resources Research*, v. 24, no. 12, p. 1997–2004.
- Helsel, D.R., and Hirsch R.M., 2002, *Statistical methods in water resources*: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chapt. A3, 510 p. (<http://water.usgs.gov/pubs/twri/twri4a3>)
- Hirsch, R.M., Slack, J.R., and Smith, R.A., 1982, *Techniques of trend analysis for monthly water quality data*: *Water Resources Research*, v. 18, no. 1, p. 107–121.
- Judge, G.G., Hill, R.C., Griffiths, W.E., Lutkepohl, H., and Lee, T.C., 1988, *Introduction to the theory and practice of econometrics* (2d ed.): New York, John Wiley, 1024 p.
- Likes, J., 1980, Variance of the MVUE for lognormal variance: *Technometrics*, v. 22, no. 2, p. 253–258.
- Powell, J.L., 1984, Least absolute deviations estimation for the censored regression model: *Journal of Econometrics*, v. 25, p. 303–325.
- Shenton, L.R., and Bowman, K.O., 1977, *Maximum likelihood estimation in small samples*: London, Charles Griffin and Co., 186 p.
- Turnbull, B.W., and Weiss, L., 1978, A likelihood ratio statistic for testing goodness of fit with randomly censored data: *Biometrics*, v. 34, p. 367–375.
- Vogel, R.M., 1986, The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypothesis: *Water Resources Research*, v. 22, no. 4, p. 587–590.

APPENDIXES

APPENDIX 1. Assignment of Detection Limits

The AMLE method described in Section 2.2.2 differs from the other load estimation methods in that a censoring threshold must be set for each and every observation in the calibration data set (a censoring threshold must be calculated even for uncensored observations). The censoring threshold is given by:

$$thresh = \ln(Q dlim) \quad (18)$$

where Q is instantaneous streamflow [L^3/T] and $dlim$ is the detection limit associated with a given observation [M/L^3]. Ideally, the detection limit would be provided by the user for each observation, such that the censoring threshold would be set at the appropriate value. Specification of a detection limit for all observations is problematic, however, in that it may place an unreasonable burden on users with large calibration data sets. Further, the detection limits associated with particular observations may be unknown for many historical data sets. Fortunately the AMLE method is relatively insensitive to the precise value of the censoring threshold when the associated observation is uncensored. Due to this fact, values of the censoring threshold within LOADEST are set based on the general procedure described below; user specification of a detection limit for every observation is not required.

General Procedure. The general procedure used to determine censoring thresholds within LOADEST is based on the approach used within ESTIMATOR. As described in Section 3.2.3, censored observations in the calibration data set are flagged using a less-than sign (<) as the first character in the CCONC field. The less-than sign is followed by the laboratory detection limit associated with the censored observation. Detection limits for the censored observations are then used to assign values of $dlim$ for the uncensored observations. These assigned values of $dlim$ are listed as part of the echo output file (echo.out, appendix 2). Assignment of $dlim$ is discussed below, where three possible cases are described in detail.

Case #1: Uncensored Data Set. The calibration data set is uncensored when all of the observations have concentrations that exceed the detection limit. When this is the case, no detection limit information is provided by the user and LOADEST assigns a $dlim$ value of 1×10^{-25} for the calculation of the censoring thresholds (equ. 18) (this arbitrary $dlim$ value is identical to that used within ESTIMATOR). The small value is assigned such that $dlim$ is less than all of the observed concentration data (that is, to reflect the fact that all of the observations are uncensored). Use of 1×10^{-25} rather than the actual detection limit is of little consequence due to the insensitivity of the AMLE method noted above. For example, data from Application 2 (Section 4.2) was reanalyzed using a $dlim$ value of 0.05 to reflect a hypothetical detection limit. Load estimates for the various time periods considered in Application 2 were relatively insensitive to the new threshold values as shown in table 26.

Table 26. Sensitivity of AMLE to value of $dlim$.

Time period	AMLE load [pounds per day]	
	$dlim=1 \times 10^{-25}$	$dlim=0.05$
Entire estimation period	21.63	21.62
Season 1	1.81	1.81
Season 2	82.03	81.97
Season 3	0.54	0.54
Season 4	1.51	1.51

Case #2: Censored Data Set with a Single *dlim*. When all of the censored observations in the calibration data set have the same detection limit, *dlim* is set equal to the value of the CCONC field of the flagged observations (*dlim* is set equal to the detection limit associated with the censored observations). This value of *dlim* then is used to calculate a censoring threshold (equ. 18) for all observations (both censored and uncensored). This calculation of the censoring threshold works perfectly when all of the observations, both censored and uncensored, have the same detection limit. Two problems arise when the calibration data set has multiple detection limits. These two problems may be illustrated using a simple example in which the calibration data set includes 14 observations from two different labs (table 27). The first seven observations have concentrations from Lab A with a detection limit of 0.01, and all of these concentrations are uncensored (CCONC > 0.01 for all observations). The latter seven observations from Lab B have a detection limit of 0.02, and one or more of the concentrations are censored. Under the general procedure described above, the detection limit for the censored observations (0.02, Lab B) is used to assign values of *dlim* for all observations, including those from Lab A. The first problem is therefore related to the calculation of the censoring threshold for all Lab A observations: the censoring threshold will be based on the assigned detection limit (Lab B, 0.02) rather than the actual detection limit (Lab A, 0.01) (see first seven observations, table 27). This problem should be of little consequence due to the insensitivity of the AMLE method noted above. A second problem is due to the lower detection limit associated with Lab A: some observations from Lab A may be above the detection limit but below the assigned value of *dlim*. This results in a contradiction in which an uncensored observation appears to be censored. The contradiction is eliminated by resetting *dlim* to 1×10^{-25} for the affected observation (see observation 4, table 27).

Table 27. Assignment of *dlim*, Case #2.

	Calibration data		Detection Limit		
	CDATE	CCONC	Actual	Assigned <i>dlim</i>	Reset <i>dlim</i>
Lab A	19781219	0.040	0.01	0.02	
	19790301	0.030	0.01	0.02	
	19790319	0.020	0.01	0.02	
	19790327	0.015	0.01	0.02	1×10^{-25}
	19790426	0.020	0.01	0.02	
	19790523	0.020	0.01	0.02	
	19790605	0.020	0.01	0.02	
Lab B	19790613	0.030	0.02	0.02	
	19790717	<0.020	0.02	0.02	
	19790815	<0.020	0.02	0.02	
	19790906	0.030	0.02	0.02	
	19790907	0.030	0.02	0.02	
	19790910	0.030	0.02	0.02	
	19790912	0.040	0.02	0.02	

Case #3: Censored Data Set with Multiple values of *dlim*. When observations within the calibration data set are censored at different levels, multiple values of *dlim* are assigned to the various observations. This assignment is illustrated in table 28, where another simple example is considered. In the example, the calibration data set includes 14 observations from two different labs. The first seven observations have concentrations from Lab A with a detection limit of 0.01; the latter seven observations are from Lab B with a detection limit of 0.02. When LOADEST is executed, the detection limit associated with the first censored observation is used to assign an initial value to *dlim*. This value of *dlim* is used in the calculation of the censoring threshold until another censored observation is encountered. For the case considered here, the first censored observation is from Lab A (see 5th observation, table 28). The detection limit from Lab A (0.01, the CCONC value preceded by a less-than sign) is therefore used to assign *dlim* for observations 1–10 (table 28). A second censored observation is then

encountered (observation 11) such that the detection limit from Lab B (0.02) is used to assign *dlim* for the remainder of the calibration data set (see observations 11–14, table 28). As shown in table 28, this procedure results in the incorrect assignment of *dlim* for observations 8–10 (the detection limit from Lab A is used to assign *dlim* for the first three observations from Lab B).

Table 28. Assignment of *dlim*, Case #3.

		Calibration data		Detection limit	
		CDATE	CCONC	Actual	Assigned <i>dlim</i>
Lab A		19781219	0.040	0.01	0.01
		19790301	0.030	0.01	0.01
		19790319	0.020	0.01	0.01
		19790327	0.015	0.01	0.01
		19790426	<0.010	0.01	0.01
		19790523	0.015	0.01	0.01
		19790605	0.020	0.01	0.01
Lab B		19790613	0.030	0.02	0.01
		19790717	0.030	0.02	0.01
		19790815	0.030	0.02	0.01
		19790906	<0.020	0.02	0.02
		19790907	0.025	0.02	0.02
		19790910	0.030	0.02	0.02
		19790912	0.040	0.02	0.02

Although the mis-assignment of *dlim* shown in this example is likely of little consequence, it should be noted that it is sometimes possible to avoid mis-assignment by reordering the calibration data set. Observations in the calibration data set are usually listed in chronological order, but reordering is permissible and will not adversely affect calibration results (reordering does, however, result in erroneous values of the “Serial Correlation of Residuals” that is reported in Part I of the constituent output file). For the case considered here, the censored observation for Lab B can be moved such that it is the first observation from Lab B within the calibration data set (table 29). This reordering results in the correct assignment of *dlim* for the entire data set (actual detection limit equals the assigned *dlim*, table 29).

Table 29. Assignment of *dlim*, Case #3 reordered.

		Calibration data		Detection Limit	
		CDATE	CCONC	Actual	Assigned <i>dlim</i>
Lab A	19781219	0.040	0.01	0.01	
	19790301	0.030	0.01	0.01	
	19790319	0.020	0.01	0.01	
	19790327	0.015	0.01	0.01	
	19790426	<0.010	0.01	0.01	
	19790523	0.015	0.01	0.01	
	19790605	0.020	0.01	0.01	
Lab B	19790906	<0.020	0.02	0.02	
	19790613	0.030	0.02	0.02	
	19790717	0.030	0.02	0.02	
	19790815	0.030	0.02	0.02	
	19790907	0.025	0.02	0.02	
	19790910	0.030	0.02	0.02	
	19790912	0.040	0.02	0.02	

APPENDIX 2. Example echo.out File from Application 4

LOADEST
A Program to Estimate Constituent Loads
U.S. Geological Survey, Version: MOD36 (Sep 2004)

White River at Hazleton IN (Station #03374100)

Echo Output File Part I: Reading from the Header File, header.inp

Estimated Values Print Option (PRTOPT): 1
Standard Error Option (SEOPT) : 3
Load Option (LDOPT) : 3

Loads for user-defined seasons are requested.

User-defined Seasons

Number of User-defined Seasons (NSEAS): 4

Season	Begin Date (SBEG)	End Date (SEND)
1	0301	0531
2	0601	0831
3	0901	1130
4	1201	0229

Model Number 0 was selected. Regression model is selected based on Akaike Information Criteria.

Number of Constituents (NCONST): 4

Constituent	Conc. Units (Flag)	Load Units (Flag)
butylate	ug/L (2)	lb/d (3)
atrazine	ug/L (2)	lb/d (3)
alachlor	ug/L (2)	lb/d (3)
suspended sediment	mg/L (1)	ton/d (4)

 Echo Output File Part II:
 Reading from the Calibration File, calib.inp

Data Variables:

Date	Time	Streamflow
19921019	1200	4.2900E+03
19921118	1145	3.1400E+04
19921221	1130	7.7700E+03
19930112	1145	3.9700E+04
19930218	1300	9.3100E+03
19930330	1400	2.0900E+04
19930412	1400	3.1300E+04
19930426	1300	1.8700E+04
19930503	1400	1.6800E+04
19930510	1500	1.8200E+04
19930517	1430	1.5600E+04
19930524	1300	9.1500E+03
19930601	1300	6.3300E+03
19930607	1500	1.0400E+04
19930614	1400	9.7600E+03
19930621	1400	1.2600E+04
19930628	1230	9.1600E+03
19930706	1400	2.1200E+04
19930712	1430	1.1100E+04
19930719	1400	1.1200E+04
19930726	1430	6.0900E+03
19930802	1430	4.5300E+03
19930809	1600	3.5600E+03
19930816	1300	6.9100E+03
19930823	1330	2.3800E+04
19930830	1330	6.7000E+03
19930907	1400	2.2700E+04
19930920	1400	7.6200E+03
19931012	1400	1.3200E+04
19931118	1150	8.7900E+04
19931123	1200	1.1300E+05
19931130	1330	3.0300E+04
19931215	1215	1.9400E+04
19940112	1245	1.1500E+04
19940215	1230	1.3600E+04
19940305	1215	1.3400E+04
19940411	1300	2.5400E+04
19940509	1330	3.3100E+04
19940523	1330	1.1700E+04
19940606	1430	5.5800E+03
19940620	1430	4.3800E+03
19940705	1330	6.3200E+03
19940718	1500	4.0600E+03
19940827	1130	1.9300E+03
19940912	1230	1.8700E+03

Constituent: butylate

Date	Time	Concentr.	Detection Limit
19921019	1200	4.0000E-03	2.0000E-03
19921118	1145	4.0000E-03	2.0000E-03
19921221	1130	2.0000E-03*	2.0000E-03
19930112	1145	9.0000E-03	2.0000E-03
19930218	1300	2.0000E-03*	2.0000E-03
19930330	1400	2.0000E-03*	2.0000E-03
19930412	1400	2.0000E-03*	2.0000E-03
19930426	1300	2.0000E-03*	2.0000E-03
19930503	1400	1.0000E-02	2.0000E-03
19930510	1500	9.0000E-03	2.0000E-03
19930517	1430	2.7000E-02	2.0000E-03
19930524	1300	4.2000E-02	2.0000E-03
19930601	1300	1.7000E-02	2.0000E-03
19930607	1500	2.4000E-02	2.0000E-03

19930614	1400	2.4000E-02	2.0000E-03
19930621	1400	6.0000E-03	2.0000E-03
19930628	1230	4.0000E-03	2.0000E-03
19930706	1400	4.0000E-03	2.0000E-03
19930712	1430	2.0000E-03*	2.0000E-03
19930719	1400	6.0000E-03	2.0000E-03
19930726	1430	2.0000E-03*	2.0000E-03
19930802	1430	2.0000E-03*	2.0000E-03
19930809	1600	2.0000E-03	2.0000E-03
19930816	1300	3.0000E-03	2.0000E-03
19930823	1330	2.0000E-03*	2.0000E-03
19930830	1330	2.0000E-03*	2.0000E-03
19930907	1400	2.0000E-03*	2.0000E-03
19930920	1400	2.0000E-03*	2.0000E-03
19931012	1400	2.0000E-03*	2.0000E-03
19931118	1150	4.0000E-03	2.0000E-03
19931123	1200	4.0000E-03	2.0000E-03
19931130	1330	2.0000E-03*	2.0000E-03
19931215	1215	3.0000E-03	2.0000E-03
19940112	1245	2.0000E-03*	2.0000E-03
19940215	1230	2.0000E-03*	2.0000E-03
19940305	1215	2.0000E-03*	2.0000E-03
19940411	1300	1.5000E-02	2.0000E-03
19940509	1330	2.6000E-02	2.0000E-03
19940523	1330	1.5000E-02	2.0000E-03
19940606	1430	2.0000E-03*	2.0000E-03
19940620	1430	2.0000E-03*	2.0000E-03
19940705	1330	2.0000E-03*	2.0000E-03
19940718	1500	2.0000E-03*	2.0000E-03
19940827	1130	2.0000E-03*	2.0000E-03
19940912	1230	2.0000E-03*	2.0000E-03

Constituent: atrazine

Date	Time	Concentr.	Detection Limit
19921019	1200	2.8000E-01	1.0000E-25
19921118	1145	3.4000E-01	1.0000E-25
19921221	1130	1.8000E-01	1.0000E-25
19930112	1145	2.3000E-01	1.0000E-25
19930218	1300	1.0000E-01	1.0000E-25
19930330	1400	1.5000E-01	1.0000E-25
19930412	1400	5.4000E-01	1.0000E-25
19930426	1300	2.0000E-01	1.0000E-25
19930503	1400	1.1000E+00	1.0000E-25
19930510	1500	1.6000E+00	1.0000E-25
19930517	1430	1.3000E+00	1.0000E-25
19930524	1300	4.7000E+00	1.0000E-25
19930601	1300	1.9000E+00	1.0000E-25
19930607	1500	9.4000E+00	1.0000E-25
19930614	1400	1.2800E+01	1.0000E-25
19930621	1400	1.0500E+01	1.0000E-25
19930628	1230	7.8000E+00	1.0000E-25
19930706	1400	3.4000E+00	1.0000E-25
19930712	1430	2.9000E+00	1.0000E-25
19930719	1400	2.1000E+00	1.0000E-25
19930726	1430	1.1000E+00	1.0000E-25
19930802	1430	9.4000E-01	1.0000E-25
19930809	1600	6.7000E-01	1.0000E-25
19930816	1300	5.9000E-01	1.0000E-25
19930823	1330	4.5000E-01	1.0000E-25
19930830	1330	8.8000E-01	1.0000E-25
19930907	1400	2.6000E-01	1.0000E-25
19930920	1400	3.1000E-01	1.0000E-25
19931012	1400	2.1000E-01	1.0000E-25
19931118	1150	9.8000E-02	1.0000E-25
19931123	1200	1.9000E-01	1.0000E-25
19931130	1330	1.7000E-01	1.0000E-25
19931215	1215	1.4000E-01	1.0000E-25
19940112	1245	9.5000E-02	1.0000E-25
19940215	1230	6.0000E-02	1.0000E-25
19940305	1215	6.5000E-02	1.0000E-25
19940411	1300	7.4000E-01	1.0000E-25
19940509	1330	5.0000E+00	1.0000E-25
19940523	1330	4.0000E+00	1.0000E-25
19940606	1430	1.5000E+00	1.0000E-25

19940620	1430	3.9000E+00	1.0000E-25
19940705	1330	6.6000E+00	1.0000E-25
19940718	1500	1.8000E+00	1.0000E-25
19940827	1130	4.1000E-01	1.0000E-25
19940912	1230	4.3000E-01	1.0000E-25

Constituent:alachlor

Date	Time	Concentr.	Detection Limit
19921019	1200	1.6000E-02	2.0000E-03
19921118	1145	4.7000E-02	2.0000E-03
19921221	1130	2.6000E-02	2.0000E-03
19930112	1145	4.7000E-02	2.0000E-03
19930218	1300	1.3000E-02	2.0000E-03
19930330	1400	3.4000E-02	2.0000E-03
19930412	1400	2.5000E-02	2.0000E-03
19930426	1300	1.8000E-02	2.0000E-03
19930503	1400	1.5000E-01	2.0000E-03
19930510	1500	7.0000E-02	2.0000E-03
19930517	1430	9.3000E-02	2.0000E-03
19930524	1300	7.4000E-01	2.0000E-03
19930601	1300	1.7000E-01	2.0000E-03
19930607	1500	5.4000E-01	2.0000E-03
19930614	1400	2.1000E+00	2.0000E-03
19930621	1400	1.2000E+00	2.0000E-03
19930628	1230	6.7000E-01	2.0000E-03
19930706	1400	2.8000E-01	2.0000E-03
19930712	1430	2.8000E-01	2.0000E-03
19930719	1400	1.2000E-01	2.0000E-03
19930726	1430	5.5000E-02	2.0000E-03
19930802	1430	3.4000E-02	2.0000E-03
19930809	1600	2.7000E-02	2.0000E-03
19930816	1300	3.1000E-02	2.0000E-03
19930823	1330	3.7000E-02	2.0000E-03
19930830	1330	5.4000E-02	2.0000E-03
19930907	1400	1.8000E-02	2.0000E-03
19930920	1400	2.0000E-03*	2.0000E-03
19931012	1400	1.9000E-02	2.0000E-03
19931118	1150	2.2000E-02	2.0000E-03
19931123	1200	4.8000E-02	2.0000E-03
19931130	1330	2.3000E-02	2.0000E-03
19931215	1215	2.1000E-02	2.0000E-03
19940112	1245	8.0000E-03	2.0000E-03
19940215	1230	1.1000E-02	2.0000E-03
19940305	1215	1.2000E-02	2.0000E-03
19940411	1300	5.8000E-02	2.0000E-03
19940509	1330	3.7000E-01	2.0000E-03
19940523	1330	3.0000E-01	2.0000E-03
19940606	1430	1.7000E-01	2.0000E-03
19940620	1430	8.5000E-02	2.0000E-03
19940705	1330	6.7000E-01	2.0000E-03
19940718	1500	7.2000E-02	2.0000E-03
19940827	1130	2.0000E-03*	2.0000E-03
19940912	1230	9.0000E-03	2.0000E-03

Constituent: suspended sediment

Date	Time	Concentr.	Detection Limit
19921019	1200	9.0500E+01	1.0000E-25
19921118	1145	1.6100E+02	1.0000E-25
19921221	1130	-9.9990E+03^	
19930112	1145	-9.9990E+03^	
19930218	1300	-9.9990E+03^	
19930330	1400	9.8000E+01	1.0000E-25
19930412	1400	1.1800E+02	1.0000E-25
19930426	1300	9.0000E+01	1.0000E-25
19930503	1400	1.3500E+02	1.0000E-25
19930510	1500	1.4000E+02	1.0000E-25
19930517	1430	1.5400E+02	1.0000E-25
19930524	1300	9.0000E+01	1.0000E-25

19930601 1300 1.2500E+02 1.0000E-25
 19930607 1500 3.6800E+02 1.0000E-25
 19930614 1400 2.5400E+02 1.0000E-25
 19930621 1400 2.3700E+02 1.0000E-25
 19930628 1230 1.5300E+02 1.0000E-25
 19930706 1400 3.4400E+02 1.0000E-25
 19930712 1430 2.2200E+02 1.0000E-25
 19930719 1400 2.6900E+02 1.0000E-25
 19930726 1430 1.2500E+02 1.0000E-25
 19930802 1430 1.1800E+02 1.0000E-25
 19930809 1600 7.3000E+01 1.0000E-25
 19930816 1300 2.7400E+02 1.0000E-25
 19930823 1330 1.9600E+02 1.0000E-25
 19930830 1330 1.2200E+02 1.0000E-25
 19930907 1400 4.9700E+02 1.0000E-25
 19930920 1400 1.2300E+02 1.0000E-25
 19931012 1400 2.5200E+02 1.0000E-25
 19931118 1150 9.5000E+01 1.0000E-25
 19931123 1200 7.5000E+01 1.0000E-25
 19931130 1330 6.3000E+01 1.0000E-25
 19931215 1215 5.3000E+01 1.0000E-25
 19940112 1245 1.0000E+01 1.0000E-25
 19940215 1230 4.3000E+01 1.0000E-25
 19940305 1215 5.0000E+01 1.0000E-25
 19940411 1300 4.7800E+02 1.0000E-25
 19940509 1330 1.1200E+02 1.0000E-25
 19940523 1330 1.9300E+02 1.0000E-25
 19940606 1430 1.1000E+02 1.0000E-25
 19940620 1430 1.3600E+02 1.0000E-25
 19940705 1330 1.0200E+02 1.0000E-25
 19940718 1500 2.0000E+01 1.0000E-25
 19940827 1130 5.3000E+01 1.0000E-25
 19940912 1230 5.3000E+01 1.0000E-25

* = Censored Observation
 ^ = Missing Observation
 # = uncensored observation is less than the detection limit that was initially assigned (assigned detection limit makes it look as if the observation is censored, when in fact it is not).
 Assigned detection limit reset to 1.E-25.

 Echo Output File Part III: Constituent Output Files

Constituent	File Name
butylate	butylate.out
atrazine	atrazine.out
alachlor	alachlor.out
suspended sediment	suspendedsediment.out

 Echo Output File Part IV:
 Reading from the Estimation File, est.inp

No. of Observations of the Data Variables per Day (NOBSPD): 1
 (The remaining lines in the Estimation File are not echoed unless an error occurs)

 Echo Output File Part V: Other Warnings and Error Messages

APPENDIX 3. Example Constituent Output File from Application 4

LOADEST
A Program to Estimate Constituent Loads
U.S. Geological Survey, Version: MOD36 (Sep 2004)

White River at Hazleton IN (Station #03374100)

Constituent: butylate

Constituent Output File Part Ia: Calibration (Load Regression)

Number of Observations : 45
Number of Uncensored Observations: 22
"center" of Decimal Time : 1993.800
"center" of Ln(Q) : 9.4824
Period of record : 1992-1994

Model Evaluation Criteria Based on AMLE Results

Model #	AIC	SPPC
1	2.602	-60.343
2	2.620	-61.652
3	2.599	-61.188
4	2.458	-58.924
5	2.632	-62.843
6	2.505	-60.870
7	2.420	-58.957
8	2.451	-60.577
9	2.487	-62.290

Akaike Information Criterion (AIC) and Schwarz Posterior Probability Criteria (SPPC) did not select same best fit model. Model # 7 selected on basis of AIC. (Model # 4 would have been selected based on SPPC)

Selected Model:

$\text{Ln}(\text{Load}) = a_0 + a_1 \text{Ln}Q + a_2 \text{Sin}(2 \text{ pi } \text{dtime}) + a_3 \text{Cos}(2 \text{ pi } \text{dtime}) + a_4 \text{dtime}$

where:

Load = constituent load [kg/d]
LnQ = Ln(Q) - center of Ln(Q)
dtime = decimal time - center of decimal time

Model Coefficients

	a0	a1	a2	a3	a4
AMLE	-2.9622	1.8007	-0.6948	-0.9079	-0.9432
MLE	-3.0019	1.8249	-0.7318	-0.9270	-0.9641

AMLE Regression Statistics

R-Squared [%] : 64.74
 Prob. Plot Corr. Coeff. (PPCC) : 0.9873
 Serial Correlation of Residuals : 0.4562

Coeff.	Std.Dev.	t-ratio	P Value
a0	0.3186	-9.30	5.444E-16
a1	0.3560	5.06	8.422E-07
a2	0.4183	-1.66	6.669E-02
a3	0.3500	-2.59	6.455E-03
a4	0.5332	-1.77	5.007E-02

Correlation Between Explanatory Variables

Explanatory variable corresponding to:

	a1	a2	a3
a2	0.5146		
a3	-0.0575	0.0990	
a4	-0.2587	-0.1901	0.0060

Additional Regression Statistics

	Residual Variance	Turnbull-Weiss Stat	DF	PL
AMLE	1.760	40.25	6	4.076E-07
MLE	2.034	31.46	6	2.071E-05

Constituent Output File Part Ib: Calibration (Concentration Regression)

AMLE Regression Statistics

Model # 7 was selected for the load regression (PART Ia) and is used here:

$$\text{Ln(Conc)} = a_0 + a_1 \text{LnQ} + a_2 \text{Sin}(\pi \text{ dtime}) + a_3 \text{Cos}(2 \pi \text{ dtime}) + a_4 \text{ dtime}$$

where:

- Conc = constituent concentration
- LnQ = Ln(Q) - center of Ln(Q)
- dtime = decimal time - center of decimal time

Concentration Regression Results

R-Squared [%] : 42.01
 Residual Variance : 1.7601

Coeff.	Value	Std.Dev.	t-ratio	P Value
a0	-6.4316	0.3186	-20.19	2.424E-24
a1	0.8007	0.3560	2.25	1.001E-02
a2	-0.6948	0.4183	-1.66	6.669E-02
a3	-0.9079	0.3500	-2.59	6.455E-03
a4	-0.9432	0.5332	-1.77	5.007E-02

 Constituent Output File Part IIa: Estimation (test for extrapolation)

Load Estimates for 19921001-19940930

Streamflow Summary Statistics [cfs]

Data	Mean	Minimum	10th Pct	25th Pct	Median	75th Pct	90th Pct	Maximum
Cal.	17632.	1870.	4198.	6515.	11500.	21050.	32080.	113000.
Est.	15788.	1280.	2804.	6378.	11100.	20225.	32170.	120000.

WARNING: The maximum estimation data set steamflow exceeds the maximum calibration data set streamflow. Load estimates require extrapolation.

Maximum Estimation Streamflow : 1.2000E+05
 Maximum Calibration Streamflow: 1.1300E+05

 Constituent Output File Part IIb: Estimation (Load Estimates)

Load Estimates for 19921001-19940930

Load Estimates [LBS/DAY]

AMLE Load Estimates

	N	Mean Load	95% Conf.Intervals		Std Error Prediction	Standard Error
			Lower	Upper		
Est. Period	730	0.68	0.22	1.62	0.37	0.35
Season 1	184	1.78	0.44	4.89	1.20	1.12
Season 2	184	0.28	0.09	0.66	0.15	0.13
Season 3	182	0.28	0.05	0.92	0.24	0.22
Season 4	180	0.36	0.09	0.97	0.24	0.22
Oct. 1992	31	2.058E-02	1.661E-03	9.025E-02	2.781E-02	2.621E-02
Nov. 1992	30	0.44	0.07	1.51	0.41	0.34
Dec. 1992	31	0.10	0.01	0.36	0.10	0.09
Jan. 1993	31	0.76	0.14	2.43	0.64	0.54
Feb. 1993	28	0.43	0.07	1.42	0.38	0.31
Mar. 1993	31	2.30	0.42	7.39	1.94	1.61
Apr. 1993	30	4.02	0.71	13.11	3.46	2.94
May 1993	31	1.58	0.34	4.70	1.19	0.92
June 1993	30	0.72	0.18	1.97	0.48	0.35
July 1993	31	0.61	0.13	1.80	0.46	0.33
Aug. 1993	31	0.22	0.04	0.73	0.19	0.12
Sep. 1993	30	0.17	0.04	0.52	0.13	0.09
Oct. 1993	31	0.11	0.02	0.30	0.07	0.06
Nov. 1993	30	0.98	0.10	4.02	1.19	1.02
Dec. 1993	31	0.21	0.04	0.67	0.17	0.14
Jan. 1994	31	0.18	0.02	0.68	0.19	0.13
Feb. 1994	28	0.52	0.08	1.81	0.49	0.39
Mar. 1994	31	0.15	0.02	0.49	0.13	0.12
Apr. 1994	30	1.58	0.18	6.16	1.77	1.48
May 1994	31	1.11	0.18	3.75	1.00	0.86
June 1994	30	0.07	0.01	0.23	0.06	0.05
July 1994	31	0.05	0.01	0.15	0.04	0.03
Aug. 1994	31	5.365E-03	4.879E-04	2.263E-02	6.802E-03	6.425E-03
Sep. 1994	30	1.901E-03	9.326E-05	9.585E-03	3.315E-03	3.188E-03

MLE Load Estimates

	N	Mean Load	Standard Error
Est. Period	730	0.79	0.38
Season 1	184	2.10	1.15
Season 2	184	0.32	0.14
Season 3	182	0.32	0.24
Season 4	180	0.40	0.24
Oct. 1992	31	2.368E-02	2.593E-02
Nov. 1992	30	0.49	0.36
Dec. 1992	31	0.11	0.09
Jan. 1993	31	0.85	0.57
Feb. 1993	28	0.48	0.33
Mar. 1993	31	2.67	1.64
Apr. 1993	30	4.83	2.85
May 1993	31	1.86	0.94
June 1993	30	0.84	0.38
July 1993	31	0.71	0.35
Aug. 1993	31	0.26	0.14
Sep. 1993	30	0.20	0.10
Oct. 1993	31	0.12	0.06
Nov. 1993	30	1.11	1.13
Dec. 1993	31	0.23	0.15
Jan. 1994	31	0.20	0.14
Feb. 1994	28	0.58	0.42
Mar. 1994	31	0.17	0.12
Apr. 1994	30	1.86	1.56
May 1994	31	1.29	0.92
June 1994	30	8.349E-02	4.806E-02
July 1994	31	5.285E-02	3.095E-02
Aug. 1994	31	6.602E-03	5.109E-03
Sep. 1994	30	2.463E-03	2.317E-03

Summary Statistics - Estimated Loads [LBS/DAY]

	Min.	25th Pct	Med.	75th Pct	90th Pct	95th Pct	99th Pct	Max.
AMLE	0.001	0.061	0.175	0.761	1.961	3.033	6.659	8.032
MLE	0.001	0.070	0.194	0.858	2.253	3.563	8.108	9.862

Summary Statistics - Estimated Concentrations [UG/L]

	Min.	25th Pct	Med.	75th Pct	90th Pct	95th Pct	99th Pct	Max.
AMLE	0.000	0.001	0.003	0.007	0.015	0.019	0.031	0.033
MLE	0.000	0.002	0.003	0.008	0.018	0.023	0.038	0.041