# USER'S GUIDE FOR
# AUTOBEALE: AN IMPLEMENTATION OF THE
# BEALE RATIO ESTIMATOR LOAD CALCULATION PROGRAM
## Featuring an Automated Optimal Stratification Search Strategy

The Beale Ratio Estimator is used to estimate mean daily loads of a water quality constituent. Required input includes a year identification code, the days of the year of interest, the mean daily flow for each of these days, and the concentrations on the days on which chemical samples were taken. The mean daily load is calculated for the days of the year for which chemical observations are available, adjusted for differences in average flow between the days on which chemical observations were made and the year as a whole, and corrected for bias which results from the correlation between flow and load. The output includes both an average daily load and a confidence interval for that load. Annual loads and their confidence intervals are obtained merely by multiplying these results by 365 (or 366 in leap years). Data for multiple calculations (i.e. different parameters) can be contained in the same data file, see later for details.

The Beale Ratio Estimator is usually used in a stratified mode. That is, parts of the year are treated separately from other parts. For example, herbicide concentrations are strongly seasonal, and it makes sense to calculate a mean daily load for the part of the year in which herbicides are present in storm runoff, and a separate mean daily load for the part of the year during which herbicides are absent or present only in low concentrations. In most applications, stratification is applied by time or by flow. However, stratification by flow is really a form of stratification in time, in which a flow criterion is used to establish the time boundaries that separate one stratum from another. The major difference is that in some forms of flow stratification, intervals of time which belong to one stratum can be interleaved with intervals of time belonging to another stratum, whereas in time stratification each stratum is continuous in time.

Because of the difficulty of identifying the best stratification scheme for a given set of data, particularly without prior experience, the present version of the program contains an algorithm that seeks to identify the optimal stratification. The criterion used is that the optimal stratification is the one that has the smallest pooled mean square error.

## Assumptions of the Beale Ratio Estimator and this implementation of it

The Beale Ratio Estimator assumes that there is a positive correlation between flux and flow (not between concentration and flow) within each stratum. This is usually approximately true of non-point pollutants. If this relationship does not hold, the ratio-based correction may be inaccurate. The program *does not* check to see that the assumed relationship holds. However, experience has shown that the Ratio Estimator is generally more robust against deviations from its assumptions than other alternatives.

The program assumes that negative concentrations have no meaning and treats all negative concentrations as missing values. This means that any missing value codes such as -1, -9, or -999 are handled correctly and need not be filtered out prior to running the program. However, if small negative concentrations are present in an uncensored data set (due to a negative value from random measurement error superimposed on a near-zero actual concentration), these values will be treated as missing. If these values should be treated as zero, they must be changed to zero in the concentration file.

Similarly, the program assumes that negative flows are not present and treats all negative values for flow as missing values. If negative flows are present due to flow reversals measured by a bi-directional current meter, the program will not calculate a negative loading component, but will treat these loads as missing, creating a positive bias in the results. There is no way to avoid this problem at present.

## Unique features of Version 3.0

Users familiar with AUTOBEALE will recognize the following differences between version 2.0 and version 3.0:

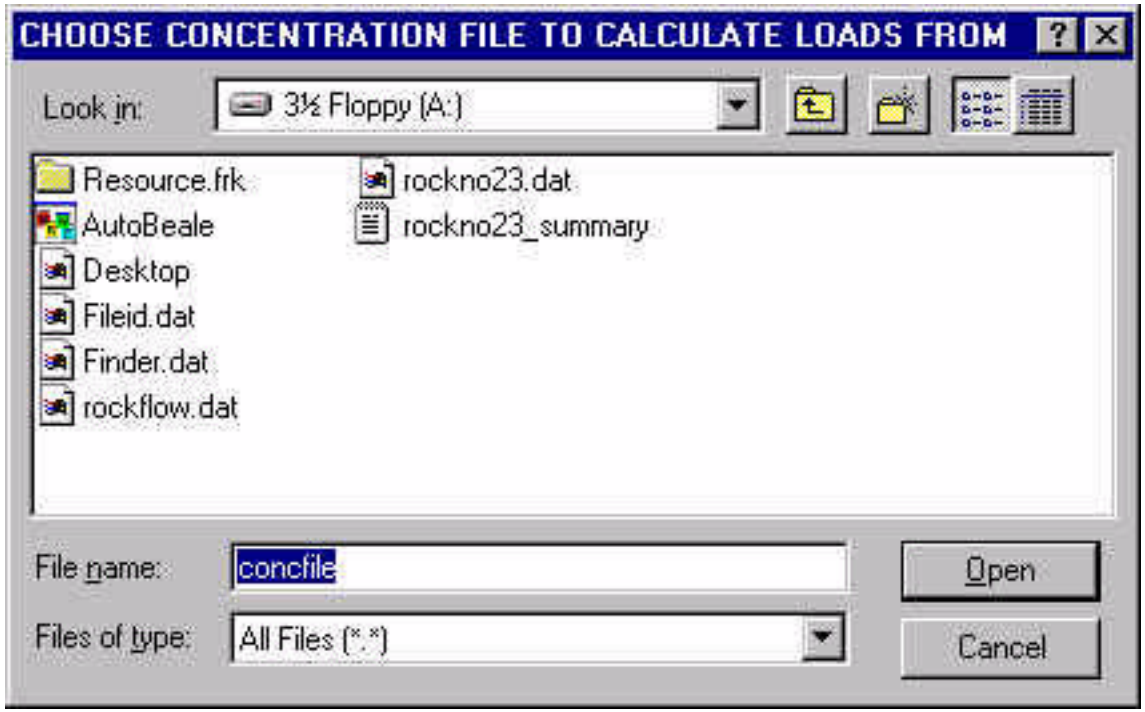
• Mean daily flows are read from one file, and concentrations are read from a second file. As a consequence, flow information need be kept only once, not once with each parameter. Considerable savings of disk space can result.

• As a further consequence of splitting the input data into chemistry and flow files, all input in a given chemistry file must be for a single station and year, because it will all use the same flow data. However, an unlimited number of chemical parameters may be run from a single file.

• Mean daily flows are used for all days. Thus, even days with concentration observations are adjusted initially so that the flow is the average discharge for the day, not the instantaneous flow at the time of the sample.

## Program Operation and Input

The program is started by double-clicking on the program icon (right). The user is first presented standard Windows file dialogs from which to choose the concentration file and the flow file (see below). **The program must be in the same folder (sub-directory) as the data files.**


The Windows file dialog

These are followed by several questions which ask the user to supply the units for concentration, flow, and the loads about to be calculated. The user may opt to save a detailed report of the load calculations as well as a standard short report (one line per load calculation).

Once this information is provided, the program proceeds to the calculation of the loads. The program operates by alternating between stratum definition and stratum adjustment until the stratification converges on the optimal one. In stratum definition mode, the program begins by placing the first stratum boundary successively on each day of the year, and recording the mean square error associated with the resulting load estimate. The day that yields the smallest mean square error is chosen as the "permanent" location of the first stratum boundary. The program then successively places the second stratum boundary on each day of the year except the one occupied by the first stratum boundary, and after comparing the results places it permanently on the day which gives the smallest mean square error, contingent upon the placement of the first stratum. This process continues until adding a further stratum on the best day either yields no further improvement in the mean square error, or yields an improvement of less than half a percent. When this point is reached, the program switches to adjustment mode.

Adjustment mode is necessary because the best possible position for the first stratum boundary may be different when the second stratum boundary is present than when it is absent, as it was when the first boundary was placed. During adjustment, each boundary is placed successively at all possible positions between the previous boundary and the succeeding boundary, and a load and mean square error are calculated and stored. If any adjusted boundary positions result in a lowered mean square error, the adjustment which lowers the mean square error the most is made permanent. Then the process is repeated until further adjustment yields no further improvement in the mean square error, or yields an improvement of less than half a percent. When this point is reached, the program switches back to stratum definition mode to see if more strata can now be added, given the adjustments to the existing strata.

When neither adjustment nor addition leads to further improvement, the final load estimate is calculated and recorded in the output files. If data for another parameter is present, the process continues.

Experience shows that adjustment usually does lead to improvement of the load estimate. After adjustment, further strata are added in about one case in four. More than two cycles of definition and adjustment are infrequently required.

## The structure of the data files (version 3.0)

Two data files are required, one containing the mean daily flows for the year, and the other containing the concentration data for which loads are to be calculated. Any reasonably short name (20 characters or less) can be used to name the input files.

Each line in the flow file contains the date in YYYYMMDD format (the millennium is coming!) and the mean daily flow (F16.4), separated by a tab. Flow may be expressed in cubic feet per second or cubic meters per second, but the units must match those chosen in the information dialog discussed above. There must be a line for each day of the year, with the day's mean daily flow. While the format F16.4 is suggested for flow, the program should read any format that includes a tab and/or one or more spaces as delimiters.

The concentration file contains an identifying name (up to 16 characters) for the data for one or more parameters covering the time defined by the flow file, the date in YYYYMMDDHHHH format (hours and minutes optional), and the concentration. Only days with concentration observations need be listed in the concentration file. If more than one concentration observation is made on a particular day, the program will average the concentrations. When the identifying name changes, the program assumes that data for a new parameter is beginning.

A load can be calculated for data of less than a year's duration, as long as the period covered contains no gaps. However, the annual load calculated from such a dataset is of questionable meaning, and requires at least careful attention to be of any worth. The annual load is calculated by multiplying the average daily load over all strata by 365. For partial years, the average daily load is probably more meaningful than this "extrapolated" annual load. Note that the period for which the load is calculated, if shorter than a year, is determined by the span of the flow data, not that of the concentration data.

A sample of a correct (flow) file follows.

```
19940501            348.0000
19940502            354.0000
19940503            360.0000
19940504            358.0000
19940505            353.0000
19940506            331.0000
```

A sample of a correct concentration file:

```
TINALA      199405021020      0.313
TINALA      199405061200      0.117
TINALA      199405091010      0.192
TINALA      199405131200      0
TINALA      199405161020      0
TINALA      199405201200      0
TINALA      199405231110      0.374
```

A warning is in order about the last lines of input files. Depending on the program used to produce these files, the last line of actual data may or may not end with a carriage return character. In particular, tab delimited files written by Excel do not contain a final carriage return. The carriage return is necessary in order for FORTRAN programs to recognize that the final line is a complete record; without it, the final line will be skipped. The easiest way to check this is to examine the input files with a word processor or editor program. When placed at the end of the file, the cursor should rest at the beginning of the line following the last data line. An alternative for Excel users is to enter something (e.g. “End”) in the first cell of the name column that follows the actual data. This will force Excel to write the next line, which will lack a carriage return and therefore be ignored by Autobeale.

Note also that if more than one empty line follows the data, this will provoke an error message when Autobeale tries to read the first empty line. Therefore, do not place unneeded empty lines at the end of the file.

## Output

Output consists of a file with a brief summary of the results and an optional detailed results. The short file file has the same name as the concentration file, but with the suffix ".txt. The detailed file, if chosen, … A sample of the printout in the detailed results file, for one stratum, is shown below.

```
TRIBUTARY NAME: MAU83ATRA
 YEAR: 1983
 PARAMETER: Maumee Atrazine sample run

 STRATUM  1
 THIS STRATUM INCLUDES DATES FROM 19830101 TO 19830502
 THE LOWER FLOW CUT OFF IS      0. CFS   THE UPPER FLOW CUT OFF IS  36900. CFS

            LOADINGS            FLOWS     CONCENTRATIONS
    DATE     KG/DAY     M3/SEC       CFS        MG/L

   19830404      6.366    744.290  26300.000       .00010
   19830411       .000    582.980  20600.000       .00000
   19830417     16.593    735.800  26000.000       .00026
   19830424      2.947    138.670   4900.000       .00025
   19830501      2.983    276.208   9760.000       .00013
   19830502     76.359    945.220  33400.000       .00094


 NUMBER OF DAYS IN THE STRATUM: 122
 MEAN STRATUM FLOW:   192.301 M3/SEC OR     6795. CFS

 MEAN SAMPLE LOADING:           17.5 KG/DAY
 MEAN SAMPLE  FLOW:   570.528 M3/SEC OR    20160. CFS

 RATIO OF MEAN STRATUM FLOW TO MEAN SAMPLE FLOW:   .34

 BIASED ESTIMATE:            5.9 KG/DAY
 UNBIASED ESTIMATE:            6.2 KG/DAY
 BIAS CORRECTION:             .3 KG/DAY

 MEAN SQUARE ERROR:         12.804 (KG/DAY)**2
 BASED ON    5. DEGREES OF FREEDOM
 SUM OF SQUARES ERROR:         76.822 (KG/DAY)**2
```

At the end of each year's report, there is a summary of the annual results, an example of which follows. Note that this is not the same stratification scheme as used in the example of a stratification file above.

```
 STRATIFICATION APPLIED:

      START    END     LOWFLOW(cfs)  HIFLOW(cfs)

  1   19830101   19830502          0.      36900.
  2   19830503   19830802          0.      10000.
  3   19830503   19830802      10000.      14000.
  4   19830503   19830802      14000.      54200.
  5   19830803   19831232          0.      39900.


 SUMMARY OVER   5 STRATA:
 MEAN DAILY LOADING:         16.337 KG/DAY
                              5963. KG/YEAR
                              5.963 TONNES/ANNUM

 MEAN SQUARE ERROR:           3.204 (KG/DAY)**2
                         426857.923 (KG/YEAR)**2
                               .427 (TONNES/ANNUM)**2

 BASED ON    18.201 DEGREES OF FREEDOM
 The 95% confidence interval half-width is          1.3720 Tonnes/Annum
```

A sample summary file is shown below. The loads are for suspended solids, total phosphorus, soluble reactive phosphorus, nitrate, and total Kjeldahl nitrogen during 1992 for Honey Creek at Melmore, Ohio.

```
   River               Load, MT/yr         MSE       DF            ± 95% CI

MEL92SS               18969.7063    8922408.7828   362.000       5854.5987
MEL92TP                  42.3887          5.1270   129.856          4.4799
MEL92SRP                  5.6045           .0062   138.420           .1551
MEL92NO3                761.2055         73.9133    39.490         17.3732
MEL92TKN                186.0787         60.6905   100.751         15.4548
```