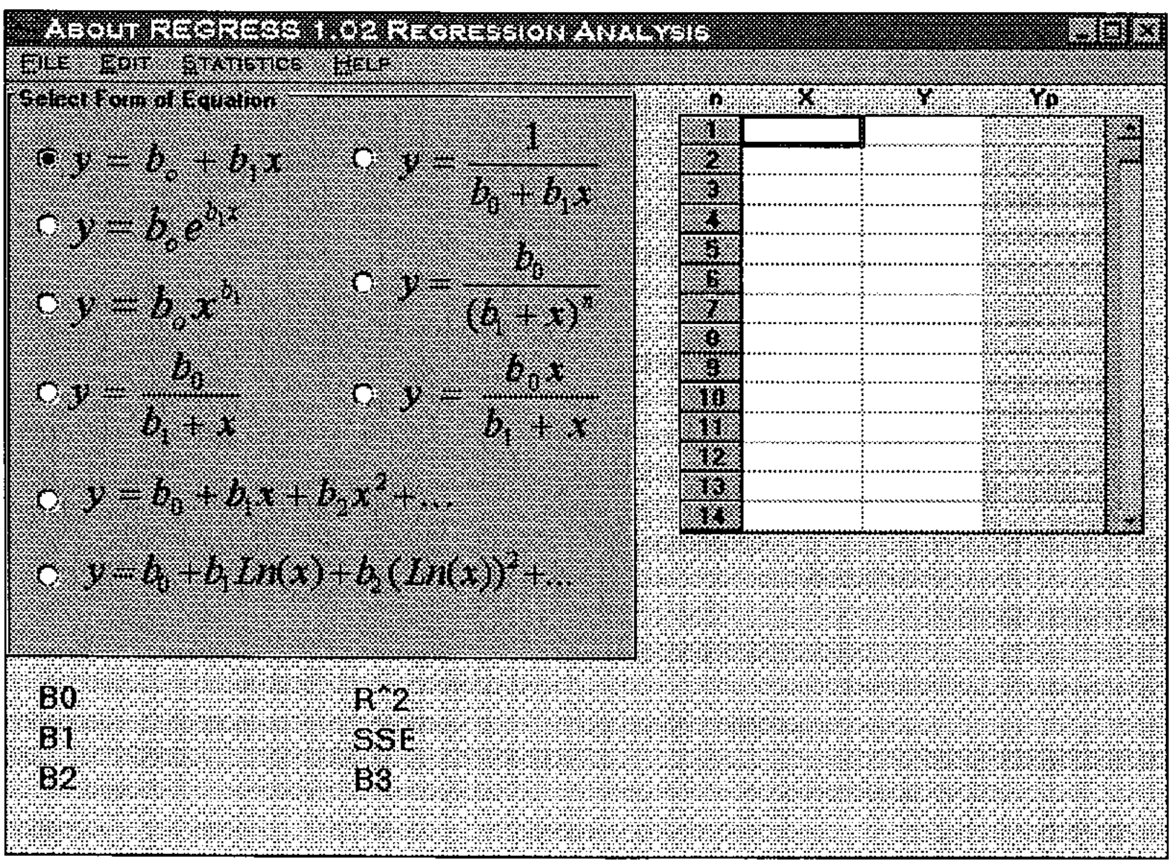
REGRESS
Statistical Regression Analysis

Introduction

REGRESS is a computer program designed to assist in performing **linear regression** analyses to assist in Hydrology. It is specifically designed to be easy to use for the hydrologist who performs regression analysis on an irregular basis and does not have time to learn the complexities of a full statistical package each time they need to perform a regression. All of the commonly required features of the program are accessible from the main window.

Using REGRESS

Regression models are located on the left side of the **About Regress 1.02 Regression Analysis** main window, with space for data entry located on the right side of the **spreadsheet**.



The regression analysis spreadsheet contains 2 columns with a white background and one column with a light blue background. The white background columns allow for the entry of **X** and **Y** (independent and dependent) data. This is the data which will be correlated in the regression analysis. Data can be directly entered into the spreadsheet or can be pasted into this spreadsheet from other Windows™ spreadsheet programs. Copy the data from the other program in columnar format. Place the cursor in the top left box (X1) of the spreadsheet in REGRESS and hit **Shift-Insert**.

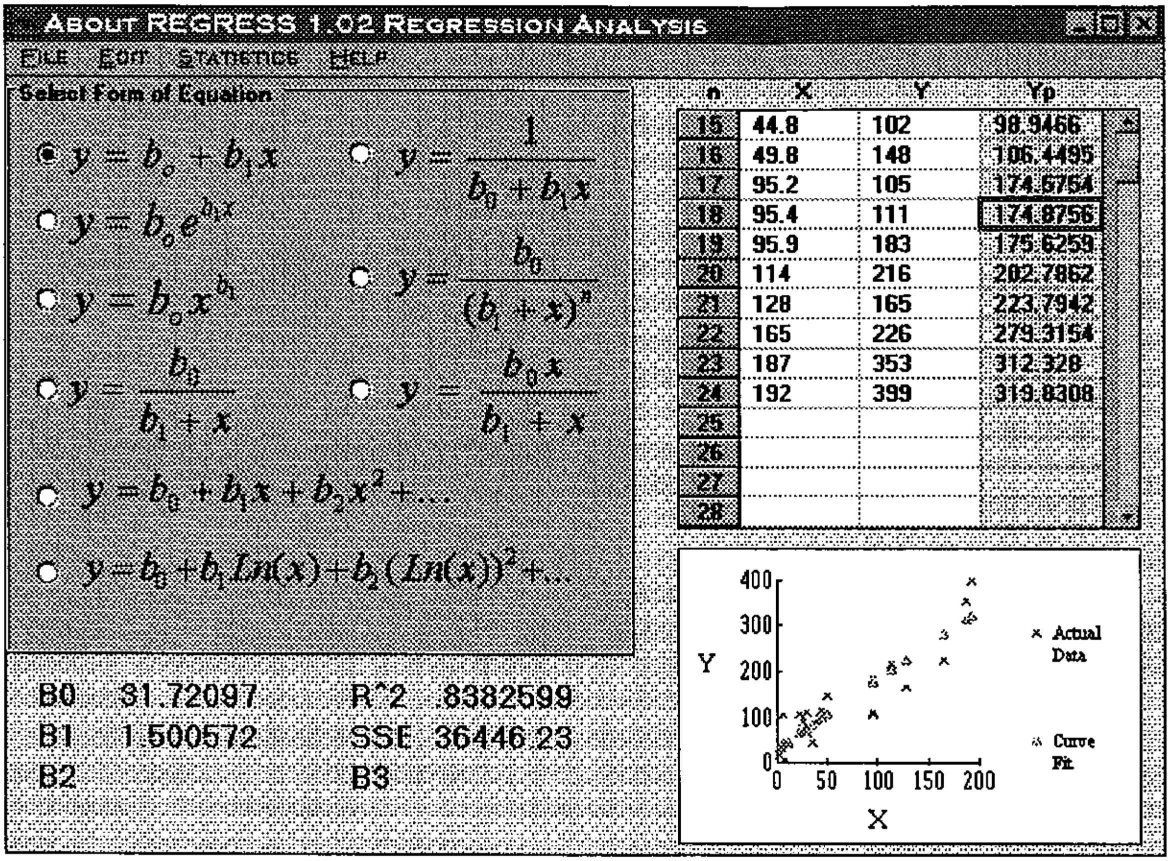
To activate the calculation process, click on the button beside the regression equation of choice in the **Select Form of Equation** list or click in the Yp column. The regression results of the fit for each of the independent variables are reported in the blue background column.

**** Regression Analysis Example Problem #1 ****

Consider the following X-Y data;

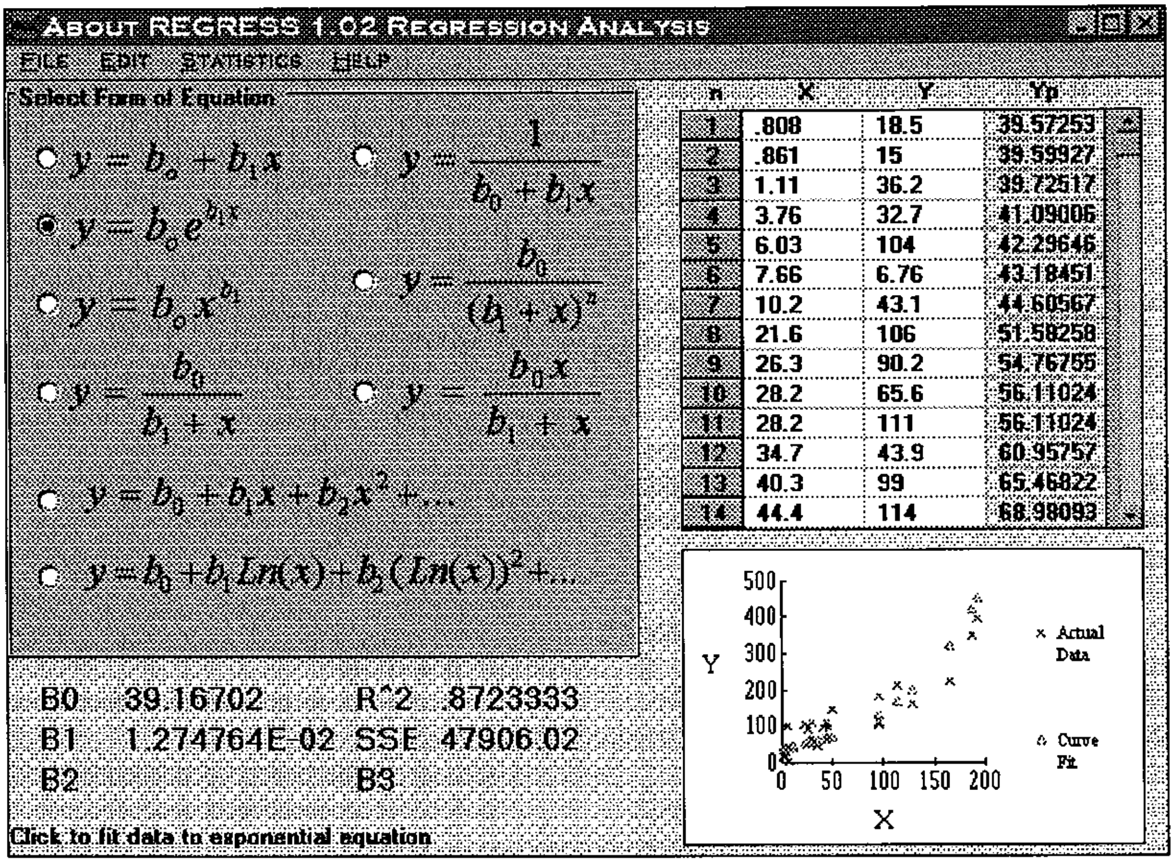
x	y	x	y	x	y
0.808	18.5	26.3	90.2	95.2	105
0.861	15	28.2	65.6	95.4	111
1.11	36.2	28.2	111	95.9	183
3.76	32.7	34.7	43.9	114	216
6.03	104	40.3	99	128	165
7.66	6.76	44.4	114	165	226
10.2	43.1	44.8	102	187	353
21.6	106	49.8	148	192	399

This data can be entered directly into the REGRESS program.



The predicted **Y** values appear in the column with the light blue background. The equation coefficients, the Sum Squared Errors **(SSE)** and the $\mathbf{R}^2$ values appear in the bottom left corner of the window.

REGRESS allows you to look at different equation forms by simply clicking the different regression models on the left side of the window.



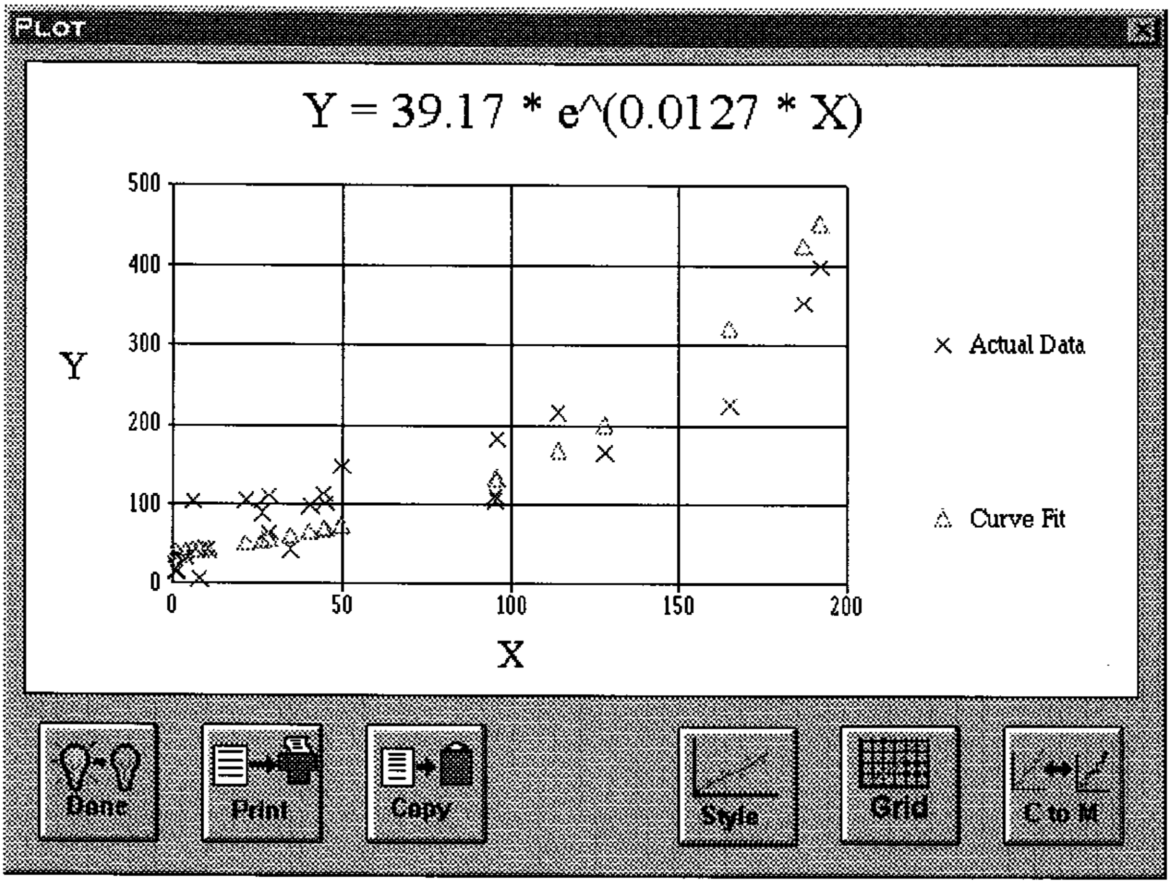
The $\mathbf{R^2}$ (correlation coefficient) value is a general measure of the 'goodness of fit' of the equation to the data.

For example, a comparison of the two equations fits shown above:

$$y = 31.72097 + 1.500572x \qquad R^2 = 0.83826$$
$$y = 39.16702e^{-0.01274764} \qquad R^2 = 0.87233$$

based on the $\mathbf{R^2}$ values, the second equation appears to be a better fit to the data.

To view the **Plot** more closely, click on the plot window and REGRESS will enlarge the plotted data.



You may now **Print** or **Copy** (to clipboard) the plot. You may also change the grid displayed by clicking on the **Grid** button. Each consecutive click will add horizontal lines, vertical lines, or return the plot to the above shown empty grid.

The "**C to M**" button switches the plot between color and monochrome display. The **Style** button allows for various plot line displays.

Select the **Done** button to return to the main REGRESS dialog window.

IDF Curve Transformation

IDF stands for Intensity Duration Frequency and is a common form used in water resources for the calculation of rainfall intensities.

The form of the IDF equation is;

$$Y = \frac{b_0}{b_1 + X}$$

This formula can be transformed to linear form through the following linear transformation.

1. First take the inverse of both sides of the equation.

$$\frac{1}{Y} = \frac{b_1}{b_0} + \frac{1}{b_0} X$$

2. By substituting Y' = 1/Y, $b'_0 = b_1/b_0$, and $b'_1 = b_0$ and we can produce the form needed for linear regression.

$$Y' = b'_0 + b'_1 X'$$

3. Regress these transformed XY values to solve for the b'_0 and b'_1 and then use these values to back calculate the original b_0 and b_1

IDF Curve Transformation Example

Using the following data perform a curve fit of the IDF form.

Duration (min)	Intensity (in/hr)	Y = 1/i
10	4.0	0.25
15	3.2	0.3125
20	2.7	0.3704
30	1.9	0.5263
60	1.2	0.8333
120	0.6	1.6667

Fit the form of equation: $i = \frac{a}{b+D}$

For linear regression we must put into the form: y = mx + b'

To convert, take inverse of both sides to yield: $\frac{1}{i} = \frac{b}{a} + \frac{D}{a}$

By setting y = 1/i, b' = b/a, m = 1/a, x = D, we can yield the proper form.

Linear least squares regression of **X** (independent) and **Y** (dependent) yields the results:

b' = b/a = 0.2303 and m = 1/a = 0.008269

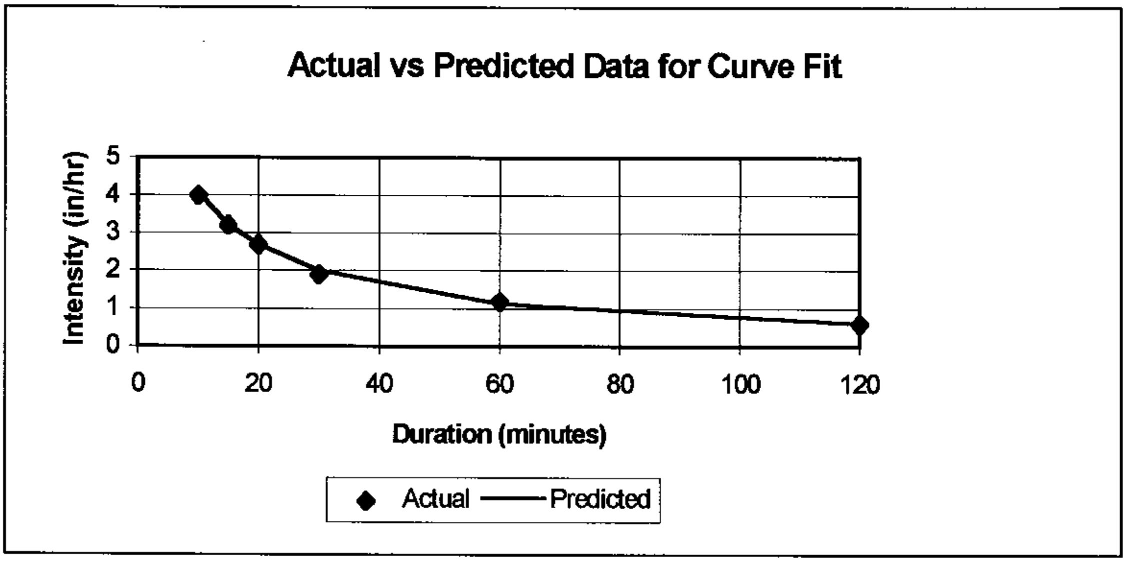
Solving for **b** and **a** gives:

a = 120.9 and b = 27.8

The form of the equation is then:

$$i = \frac{78.38}{9.22 + D}$$ with **i** in in/hr and **D** in minutes.

A plot of the predicted and actual results is:



Monod Transformation

The Monod is a common equation used in biological processes for the calculation of cellular growth. The form of the Monod equation is;

$$Y = \frac{b_0 X}{b_1 + X}$$

This formula can be transformed to linear form through the following linear transformations.

1. Lineweaver - Burke Transformation

$$\frac{1}{Y} = \frac{b_1}{b_0} X + \frac{1}{b_0}$$

2. Haines Transformation

$$\frac{X}{Y} = \frac{b_1}{b_0} + \frac{1}{b_0} X$$

3. Edy-Hoffstead

$$Y = b_0 - b_1 \frac{Y}{X}$$

These three expressions fit the linear form and are mathematically equivalent to the original form of the Monod equation.