

IJP 02585

## Research Papers

# Mathematical transformation of data, fit, and graphic representation of regression curves through computer programming

Manuel Veiga and Filomena Gutiérrez

*Departamento de Microbiología, Facultad de Farmacia, Universidad de Santiago, 15706 Santiago de Compostela (Spain)*

(Received 24 January 1991)

(Accepted 1 July 1991)

**Key words:** BASIC; Computer program; Statistics; Regression curve; Plotting; Mathematical transformation

---

### Summary

A computer program for the graphic representation of regression curves from two-dimensional data is presented. The program, written in BASIC, performs the fit and the graphic display of linear, exponential, geometric and polynomial curves. The program also performs mathematical manipulations of data producing graphic displays from the transformed data, as well as the calculation of several statistics (coefficient of correlation, coefficient of determination, sum of squares of the deviations and standard error of the estimate) for the different regression curves.

---

### Introduction

The use of computers in scientific laboratories has increased significantly during recent years and computers are widely used in the everyday analysis of biomedical laboratory data. Most research workers like to see their data in graphical form. A simple perusal of the scientific literature shows that one of the most common ways of examining the relationship between two variables is the plotting of one variable vs another on a two-dimensional graphic. Such graphics constitute an effective means for presenting information concisely and clearly in final reports; they are also useful in assessing the fit of the data to

known theoretical models (such as regression curves) and a very important aid in interpreting the data analysis.

In this paper, we describe a computer program which produces graphical displays of two-dimensional data and performs mathematical transformations of data, and thus is very useful in biomedical research. It is especially intended to help researchers working at these scientific laboratories.

### The Program

The program \*, which was developed on an IBM AT computer, has been written in GWBA-

---

*Correspondence:* M. Veiga, Departamento de Microbiología, Facultad de Farmacia, Universidad de Santiago, 15706 Santiago de Compostela, Spain.

---

\* Readers who would like to obtain a copy of the computer program are requested to contact the author.

SIC using a very simple menu-driven system in which the user is guided through the options at any given stage. To help in understanding the program, the headlines of the different subroutines inform the user of their aim. Some of the program subroutines are based upon those developed in a previous program (Veiga and Gutiérrez, 1991).

One of the advantages of this program relies on the fact that the user has only to enter the scientific data following the instructions shown on the screen and the data will be saved in a file and, within a few seconds, a straight line fitted by the least-square method will be displayed on the screen. At this stage, the user, by simply pressing a key, can fit another curve to the data (exponential, geometric or polynomial), perform a number of manipulations of the data (the computer will draw a curve fitted to the new points) or assess the fit of the data to the different regression curves.

Once the user has entered his/her data, the program itself creates the corresponding scales for the axes, adjusting them to their optimal values for a good representation. If the scales of the axes are unsatisfactory, they can easily be changed by entering the desired values directly from the keyboard. The adjustment of the plotting scales allows one to zoom in on an area of interest. The user can obtain a hard copy of the graphic by pressing PrScr on the keyboard, and if an adequate word processing program is available, the screen can be captured and included in a *File.doc* to be read and printed at any time. To improve the graphic resolution, when working with a computer with a VGA monitor, the current screen ( $640 \times 200$  pixels) can be changed into a VGA one ( $640 \times 480$  pixels) by pressing a key. In this case, since GWBASIC does not allow this high-resolution graphic, the program must be run by using an alternative and BASIC-compatible language (Turbobasic etc.). If other monitors are available, such as EGA, etc., the graphic resolution may be improved by modifying the values of S, Q, Y1, Y2 and WDW at program lines 31 and 29000–29020. The program has several other important features. Table 1 lists the different available options after plotting a graphic.

TABLE 1

Main options

Option	Purpose
Q	Returns to the main menu (Quit)
L (R)	Fits a Linear curve to the data
E	Fits an Exponential curve to the data
G	Fits a Geometric curve to the data
P	Fits a Polynomial curve to the data <sup>a</sup>
N	Draws the points fitting No curve
F	Draws fitted curve From $X_1$ to $X_n$ <sup>b</sup>
M	Allows Mathematical transformations of the data <sup>c</sup>
A	Shows the Accuracy of the fit
D	Draws axes passing through (0, 0)
C	Allows one to Change the scales of the graphic axes <sup>d</sup>
S	Changes the Screen resolution
Y	Allows extrapolation of the $X$ values
X	Allows extrapolation of the $Y$ values

<sup>a</sup> Once this option has been chosen, the computer prompts to enter the required degree of the polynomial.

<sup>b</sup> If this option is not chosen, the curve will be drawn along the entire  $XX$  axis.

<sup>c</sup> The mathematical manipulations available are: Lineweaver-Burk plot; Eadie-Hofstee plot; logarithmic plots; user-defined conversions.

<sup>d</sup> If <Enter> is pressed when the computer prompts to enter the highest and lowest values of each axis, the initial values for this axis will remain unchanged.

<sup>e</sup> Graphics obtained from options N, F, D and S are reversed by pressing the corresponding key again.

## Computations

All computations carried out by this program are performed with *double precision*. All curves are fitted by the method of least squares. The available curves in this program are the following:

Linear curve:	$Y = a + bX$
Exponential curve:	$Y = ab^X$
Geometric curve:	$Y = aX^b$
Polynomial curve of degree $n$ :	$Y = a_0 + a_1X + a_2X^2 + \dots a_nX^n$

When the program is run, the computer will fit the data to a linear, an exponential and a geometric curve and the corresponding values for the independent term ( $a$ ), the curve slope ( $b$ ) and the coefficients of correlation ( $R$ ) will be calculated. If the user chooses to fit a polynomial curve to

his/her data, the computer will ask him/her to enter the degree of the polynomial. Once the degree of the polynomial has been entered, the computer proceeds to set up a matrix of simultaneous equations. There is one equation for each term in the final curve. The matrix is solved by Gauss' method and the coefficients  $a_0 \dots a_n$  calculated.

The accuracy of the fit of the different curves to the data is estimated by calculating the values of the coefficients of correlation ( $R$ ) and determination ( $R^2$ ) and the standard error of the estimate (SE). After fitting the corresponding curves, the standard errors of the estimate are calculated from:

$$SE = \sqrt{\Sigma D^2 / N}$$

where  $N$  is the number of observations and  $\Sigma D^2$  (residual sum of squares) represents the sum of the squares of the deviations of observations from their corresponding calculated values:

$$\Sigma D^2 = \Sigma (\hat{Y}_i - Y_i)^2$$

The value of  $R^2$  is estimated using:

$$R^2 = \frac{\Sigma (\bar{Y} - \hat{Y}_i)^2}{\Sigma (\bar{Y} - Y_i)^2}$$

where  $\Sigma (\bar{Y} - Y_i)^2$  is the sum of the squares of the deviations of observations from their mean and  $\Sigma (\bar{Y} - \hat{Y}_i)^2$  is the sum of the squares of the deviations of the calculated values of  $Y$  from the mean of  $Y$ .

## Mathematical Manipulations

Simple arithmetic calculations are frequently useful during analyses of biomedical data for a variety of reasons. One of the primary uses of simple calculations is for transforming data prior to analysis. There are many transformations and the reasons for using them depend on the data or analysis. Most of the time, such operations are

time-consuming and it is most advantageous to perform them directly without leaving the program. All mathematical transformations described in the present article are carried out by the program itself. To use this option, the data must be previously saved in a file. The available options on the mathematical manipulations menu are the following:

Lineweaver-Burk plot

Eadie-Hofstee plot

Semilogarithmic conversion [ $Y \rightarrow \log Y$ ]

Semilogarithmic conversion [ $X \rightarrow \log X$ ]

Double-logarithmic conversion

[ $Y \rightarrow \log Y$ ][ $X \rightarrow \log X$ ]

User-defined conversion

As described earlier, when the dependent variable ( $Y$ ) has a nonlinear relationship with the independent variable ( $X$ ), the user can choose to fit his/her data to the other available curves. The nonlinear exponential ( $Y = ab^X$ ) and geometric ( $Y = aX^b$ ) equations can be converted into linear ones by taking logarithms.

$$Y = ab^X \Rightarrow \log Y = \log a + \log b(X)$$

$$Y = aX^b \Rightarrow \log Y = \log a + b(\log X)$$

therefore, plotting  $\log Y$  against  $X$  (exponential) or  $\log Y$  against  $\log X$  (geometric) results in straight lines. Thus, if the data fits to an exponential or a geometric curve, the user can choose a semilogarithmic conversion (exponential) or a double-logarithmic conversion (geometric) on the mathematical manipulations menu and the resulting line will be plotted superimposed on the transformed data points. Semilogarithmic and double-logarithmic representations can be of considerable advantage for a rapid examination of the behaviour of different parameters (microbial growth, mutagenic activity, antibiotic activity, etc.).

Other mathematical transformations of the data are those which lead to the Lineweaver-Burk and Eadie-Hofstee plots. They are widely employed in the study of enzyme kinetics, since they allow the values of the Michaelis constant ( $K_m$ )

and enzymatic velocities to be determined very readily.

In the Lineweaver-Burk plot, the values of  $X$  and  $Y$  are transformed into their reciprocals,  $1/X$  and  $1/Y$ . The maximum velocity and the Michaelis constant are given, respectively, by  $V_{\max} = 1/a$  and  $K_m = b/a$ . In the Eadie-Hofstee plot, the  $Y$  remain unchanged and the  $X$  are converted into  $Y/X$ . The corresponding values for  $V_{\max}$  and  $K_m$  are  $V_{\max} = a$  and  $K_m = -b$ . The Eadie-Hofstee plot not only yields the values of  $V_{\max}$  and  $K_m$  in a very simple manner, but is also presents on a greatly enlarged scale deviations of

the data as a linear representation, which is beyond the scope of the Lineweaver-Burk plot.

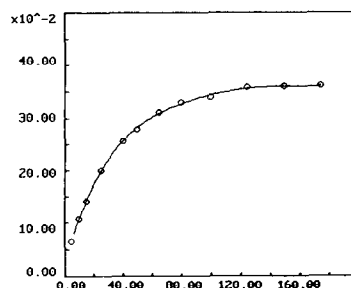
The user-defined conversion allows any transformation of the data, provided that such a transformation can be written in BASIC. Once this option has been chosen, the computer will prompt  $X = ?$ , and then  $Y = ?$ , demanding to write the equations in BASIC. To illustrate this function, imagine that the  $X$  represents temperature (in  $^{\circ}\text{C}$ ) and that the  $Y$  denotes the decomposition rate of a drug ( $K_{\text{obs}}$ ) (in  $\text{h}^{-1}$ ). The user could choose, for instance, an Arrhenius plot of the logarithm of the observed rate of decomposition

Fig. 1.- Cellulase activity from Streptomyces M54.

Number of points,  $N = 12$

$a = 168.7847511373766$   
 $b = 103.3517053423813$   
 $c = -1.494996116408206$   
 $d = 1.220338563771623\text{E}-002$   
 $e = -5.206801002630997\text{E}-005$   
 $f = 8.876761824595977\text{E}-008$

$R = 0.9999$



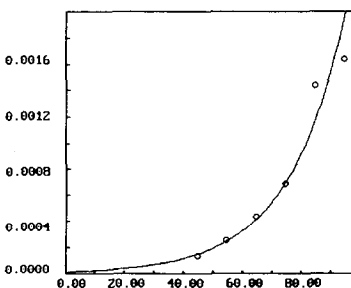
$$Y = a + bX + cX^2 + dX^3 + eX^4 + fX^5$$

Fig. 3.- Decomposition of drug D11 as a function of temperature

$X = ^{\circ}\text{C}$   
 $Y = K_{\text{obs}} (\text{h}^{-1})$

X values	Y values
45.00	0.0001
55.00	0.0003
65.00	0.0004
75.00	0.0007
85.00	0.0014
95.00	0.0016

Number of points,  $N = 6$



$$Y = 0.000014 + 1.053300023078918 \cdot X$$

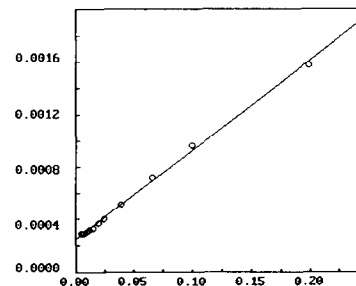
$R = 0.9909$

Fig. 2.- Cellulase activity from Streptomyces M54. Lineweaver-Burk.

$X = 1/(\text{min})$   
 $Y = 1/(U/\text{ml})$

X values	Y values
0.2000	0.0015
0.1000	0.0009
0.0667	0.0007
0.0400	0.0005
0.0250	0.0004
0.0200	0.0004
0.0154	0.0003
0.0125	0.0003
0.0100	0.0003
0.0069	0.0003
0.0067	0.0003
0.0057	0.0003

Number of points,  $N = 12$



$$Y = 0.000232 + 0.006669 X$$

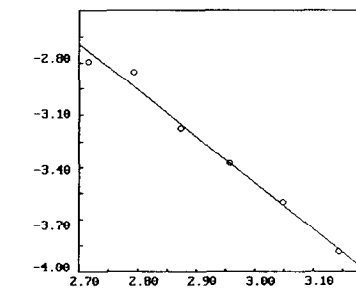
$R = 0.9988$

Fig. 4.- Decomposition of drug D11 as a function of T. Arrhenius plot.

$X = 1000/(X+273)$   
 $Y = \log(Y)/\log(10)$

X values	Y values
3.14	-3.89
3.05	-3.60
2.96	-3.38
2.87	-3.17
2.79	-2.85
2.72	-2.88

Number of points,  $N = 6$



$$Y = 4.458700180053711 - 2.648263931274414 X$$

$R = -0.9936$

Fig. 1. Example of a typical plot. The data was fitted by the least-square method to a polynomial curve of degree 5. The data is fictitious.

Fig. 2. The data from Fig. 1 was mathematically transformed to obtain a Lineweaver-Burk plot, and fitted by the least-square method to a straight line. The data is fictitious.

Fig. 3. Set of data fitted by the least-square method to an exponential curve. The data is fictitious.

Fig. 4. The data from Fig. 2 was mathematically transformed to obtain an Arrhenius plot, and fitted by the least-square method to a straight line. The data is fictitious.

as a function of the reciprocal of the absolute temperature by writing the following equations:

$$X = ? 1000 / (X + 273)$$

$$Y = ? \log(Y) / \log(10)$$

All the  $X$  would be converted into  $1/T \times 10^3$  (K), and the  $Y$  into  $\log K_{\text{obs}}$  ( $\text{h}^{-1}$ ).

This function cannot work in compiled languages such as Turbobasic, since it is performed through a modification of the program itself by substitution of program line 4050. If the user had

chosen the previously mentioned transformation, line 4050 would appear as follows:

$$4050 \text{ DEF FNX}(X, Y) = 1000 / (X + 273):$$

$$\text{DEF FNY}(X, Y) = \text{LOG}(Y) / \text{LOG}(10)$$

Finally, it should be pointed out that the program is provided with an error code subroutine to detect any wrongly conducted manipulation of the data or of the program itself.

Figs 1-4 illustrate four different graphics

Fig. 5.- Cellulase activity from Streptomyces MS4.

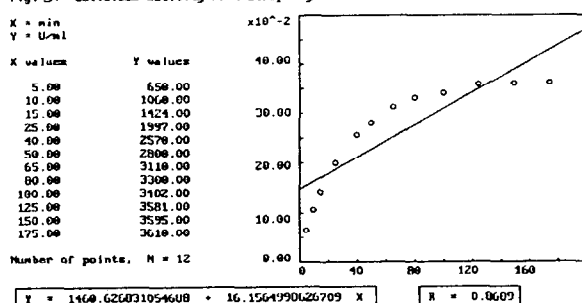


Fig. 6.- Cellulase activity from Streptomyces MS4.

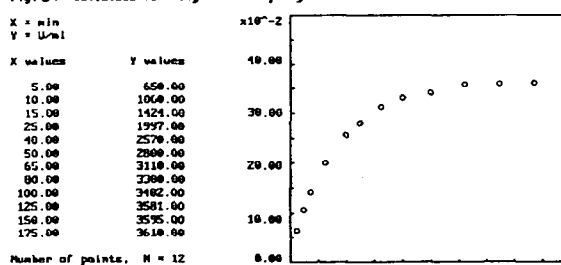


Fig. 7.- Cellulase activity from Streptomyces MS4.

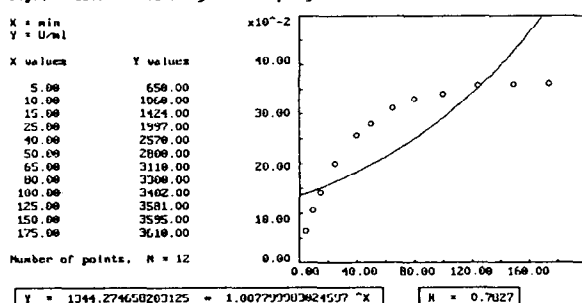


Fig. 8.- Cellulase activity from Streptomyces MS4.

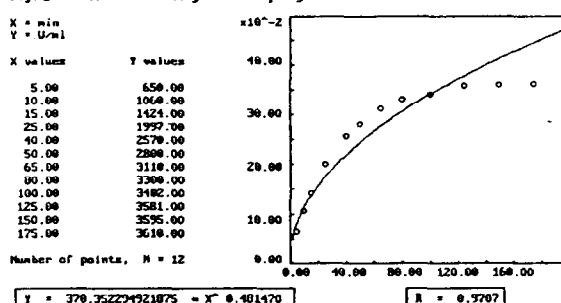


Fig. 9.- Cellulase activity from Streptomyces MS4.

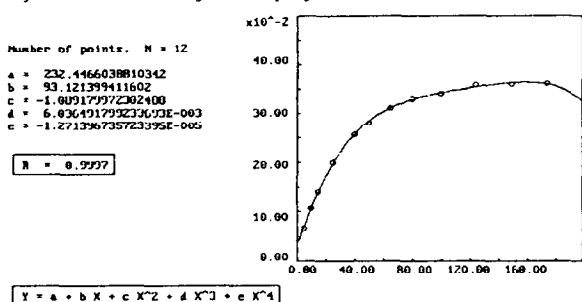


Fig. 10.- Cellulase activity from Streptomyces MS4.

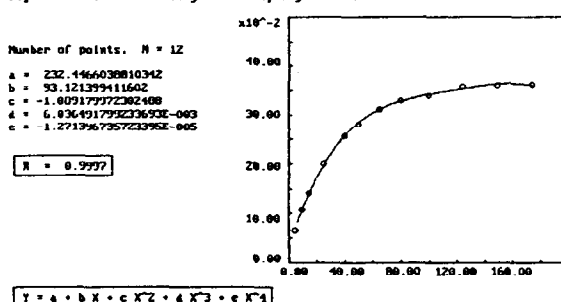


Fig. 5. Option L. Fig. 6. Option N. Fig. 7. Option E. Fig. 8. Option G. Fig. 9. Option P-4. Fig. 10. Option F.

Fig.11.- Cellulase activity from Streptomyces NS4.

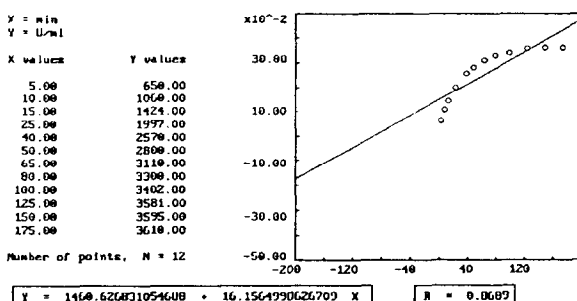


Fig.12.- Cellulase activity from Streptomyces NS4.

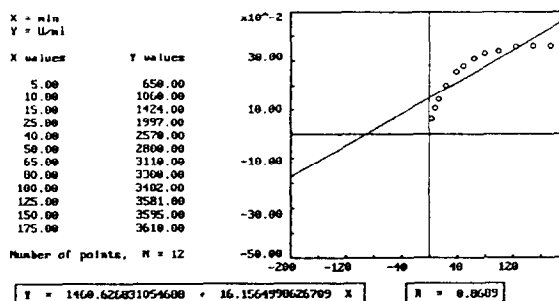


Fig.13.- Cellulase activity from Streptomyces NS4.

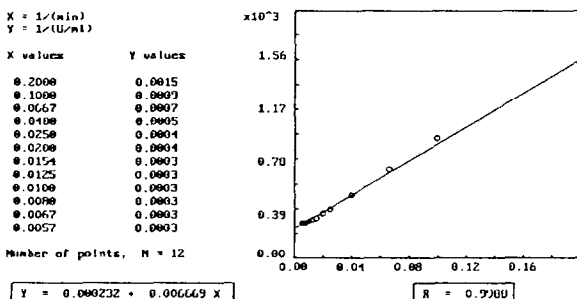


Fig.14.- Cellulase activity from Streptomyces NS4.

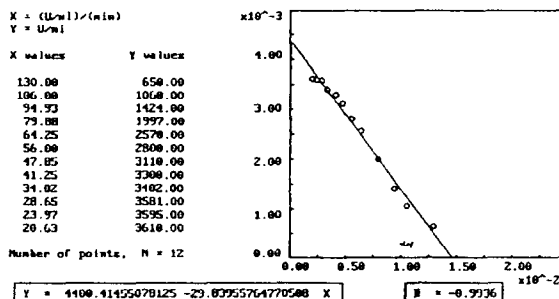


Fig.15.- Cellulase activity from Streptomyces NS4.

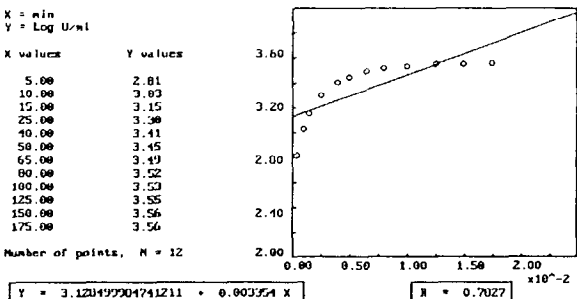


Fig.16.- Cellulase activity from Streptomyces NS4.

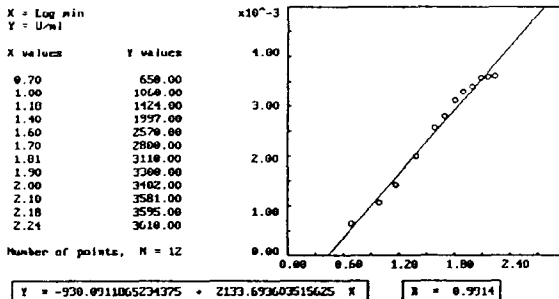


Fig.17.- Cellulase activity from Streptomyces NS4.

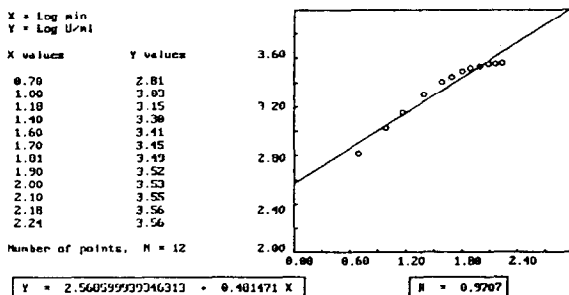


Fig.18.- Cellulase activity from Streptomyces NS4.

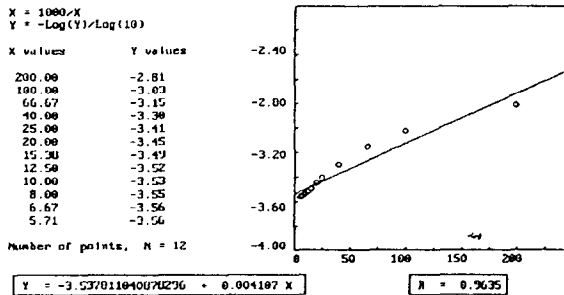


Fig. 11. Option C. Fig. 12. Option D. Fig. 13. Option M-1. Fig. 14. Option M-2. Fig. 15. Option M-3. Fig. 16. Option M-4.

Fig. 17. Option M-5. Fig. 18. Option M-6.

drawn after running the program. The four graphics were obtained from two sets of data.

## References

Veiga, M. and Gutiérrez, F., Basic computer program for the graphic representation of microbial growth curves and batch fermentations. *J. Microbiol. Methods*, 13 (1991) 23–28.

## Acknowledgement

This study was partially supported by grant XUGA8151589 from Xunta de Galicia.

## Operating Instructions

### Introduction

- (1) The program was written to be used with any IBM compatible computer. The data diskette must be loaded in drive A. The program itself may be installed in any unit (A, B, C, etc.).
- (2) The program was designed to work in Screen mode 2 (GWBasic, CGA monitor), Screen mode 9 (GWBasic, EGA monitor) and Screen mode 11 (TurboBasic, VGA monitor). If other monitors are available, the values of S, Q, Y1, Y2 and WDW must be modified. If option S does not work, the user can try to modify these values.
- (3) Option 6 of the mathematical menu cannot work in compiled programs such as those generated by TurboBasic.
- (4) In the figures shown in this manual, bold letters correspond to legends produced by the program. Normal letters correspond to entries typed by the user.
- (5) Any option in this program is selected by pressing the corresponding key.
- (6) All figures shown below were produced by the program from the data shown on page 8.
- (7) The files created by this program are termed with the extension .RTA. Each time the user wishes to open a file, the computer shows all files with the extension .RTA in drive A as follows:

### DIRECTORY

```
A:\
COLI      .RTA    EMS      .RTA    MNNG     .RTA
```

```
1209344 Bytes free
```

```
Enter FILENAME (w/o extension) ?
```

As can be seen, the data diskette in unit A has three files generated by this program. These files can be read at any time. To open a file the user has to enter the corresponding filename omitting the extension.

### Running the Program

When the program is run, the following menu appears on the computer screen:

```
SELECT OPTION
```

- ```
1.- DRAW GRAPHIC FROM A DATA FILE
2.- ENTER AND SAVE DATA
3.- ENTER DATA w/o SAVING
4.- MATHEMATICAL MANIPULATIONS
5.- EXTRAPOLATION
6.- ACCURACY OF THE FIT FOR DIFFERENT CURVES
```

```
PRESS <ESC> TO END PROGRAM
```

At this stage, options 1, 4, 5 and 6 can only be used with data previously saved in a file. Therefore, when running the program for the first time, options 2 or 3 must be chosen.

If option 2 is chosen, the following will appear consecutively on the screen:

```
Enter NAME for this FILE ?
```

```
Enter a REMARK (e.g. Fig. 5.- Glucose standard curve)
?
```

```
XX units ?
YY units ?
Number of points ?
```

The user has to enter a filename and the number of points. The other entries can be typed or left without filling. An example is set out below:

```
Enter NAME for this FILE ? CELL-N54
```

```
Enter a REMARK (e.g. Fig. 5.- Glucose standard curve)
? Fig. 1.- Cellulase activity from Streptomyces N54.
```

```
XX units ? min
YY units ? U/ml
Number of points ? 12
```

The program will then ask the user to enter his/her data, first the  $X$  values and then the  $Y$  values:

| Register No. | X values | Y values |
|--------------|----------|----------|
| 1.-          | 5        | 650      |
| 2.-          | 10       | 1060     |
| 3.-          | 15       | 1424     |
| 4.-          | 25       | 1997     |
| 5.-          | 40       | 2570     |
| 6.-          | 50       | 2800     |
| 7.-          | 65       | 3110     |
| 8.-          | 80       | 3300     |
| 9.-          | 100      | 3402     |
| 10.-         | 125      | 3581     |
| 11.-         | 150      | 3595     |
| 12.-         | 175      | 3610     |

ANY CORRECTION ? (y/n) N

Once this operation has concluded, the program saves the data in a file named CELL-N54.RTA and displays a straight line fitted to the data (Fig. 5). At this stage the user can select any option shown in Table 1. Figs 6–10 show, respectively: (6) no curve fitted to the data (option N); (7) exponential curve (option E); (8) geometric curve (option G); (9) polynomial curve of degree 4 (option P); (10) curve drawn from  $X_1$  to  $X_n$  (option F).

If the user wishes to change the graphic scales, option C should be selected and the desired values will be entered as follows:

```

GRAPHIC AXES SCALES
PRESENT VALUES: X max=200 Y max=5000
                  X min=0   Y min=0

Maximum value for XX axis ? 200
Minimum value for XX axis ? -200

Maximum value for YY axis ? 5000
Minimum value for YY axis ? -5000

```

Once these new values have been entered, the program produces a graphic such as that shown in Fig. 11.

If option D is chosen, then the program will produce the graphic shown in Fig. 12.

### Mathematical Transformation of Data

If option M is chosen (or option 4 in the main menu), the mathematical transformation menu will appear on the screen as follows:

```

SELECT OPTION
1.- Lineweaver-Burk plot

```

```

2.- Eadie-Hofstee plot
3.- Semilogarithmic conversion [YY=Log YY]
4.- Semilogarithmic conversion [XX=Log XX]
5.- Logarithmic conversion [XX=Log XX, YY=
   Log YY]
6.- User defined conversion

```

The plots produced after any mathematical transformation of the data can be saved into new files. Figs 13–17 show, respectively, the plots corresponding to options 1–5 on this menu. If option 6 is chosen the user will have to enter his/her transformations as in the following example:

```

WRITE THE EQUATIONS IN BASIC

X=? 1000/X
Y=? -LOG(Y)/LOG(10)

```

and the program will produce the plot shown in Fig. 18.

### Extrapolations

This option can be chosen at the main menu (option 5) or at any time while the program is being run. It works after fitting any sort of the available curves. Thus, if a straight line is fitted to the data, when option Y is chosen, the following will appear consecutively on the computer screen:

```

How many POINTS do you want to calculate ? 3

X range: 0...200
Y range: 0...5000

INPUT X values

? 10
? 20
? 55

```

and then the following:

```

Y=1460.627+16.1565 X R=0.8689

X units: min
Y units: U/ml

X values      Y values
10            1622.192
20            1783.757
55            2349.235

```

PRESS ANY KEY TO CONTINUE...

In cases where the  $Y$  units are given in logarithms, the program itself will calculate the corresponding antilogarithms.



The following example shows data extrapolation after semilogarithmic transformation of the data (mathematical menu, option 3) has been carried out:

$$Y=3.1285+0.0034 \times X \quad R=0.7827$$

X units: min

Y units: Log U/ml

| X values | Y values | Antilog Y |
|----------|----------|-----------|
| 10       | 3.16203  | 1452.212  |
| 20       | 3.195573 | 1568.817  |
| 55       | 3.312971 | 2055.753  |

PRESS ANY KEY TO CONTINUE...

### Accuracy of the Fit

If option A is chosen (or option 6 at the main menu), the program will produce the following:

|             | R      | R <sup>2</sup> | $\sum D^2$ | Std. ERROR<br>of est. |
|-------------|--------|----------------|------------|-----------------------|
| Linear      | 0.8689 | 0.7549         | 3020692    | 501.7213              |
| Exponential | 0.7828 | 0.6127         | 6741316    | 749.5175              |
| Geometric   | 0.9707 | 0.9423         | 1595784    | 364.6669              |
| Polynomial  |        |                |            |                       |
| Degree 2 .- | 0.9839 | 0.9681         | 392743.1   | 180.9105              |
| Degree 3 .- | 0.9979 | 0.9959         | 50557.84   | 64.9088               |
| Degree 4 .- | 0.9997 | 0.9995         | 6968.991   | 24.0987               |
| Degree 5 .- | 0.9999 | 0.9998         | 3437.437   | 16.9249               |

PRESS ANY KEY TO CONTINUE...