An Open Source Modelling Software for Scientific and Engineering Applications

O. Onuoha, Prof. H.C. Inyiama, U. Onochojah, Dr. C.C. Okezie, L. Ezema, M.U. Abuh

Projects Development Institute (PRODA)

Onuohaokechi2000@yahoo.com

Nnamdi Azikiwe University (NAU), Awka

drhcinyiama@gmail.com

Projects Development Institute (PRODA)

ufuomaon@yahoo.com

Nnamdi Azikiwe University (NAU), Awka

Federal University of Technology, Owerri (FUTO)

ezemms@yahoo.com

University of Port Harcourt

abuhnature@yahoo.com

ABSTRACT

Modelling commonly entails the establishment of relationship(s) between data variables. One common way of doing this, after the data to be modelled has been collected, is to plot a scatter diagram of the data samples collected from the field and then proceed to draw the graph and then derive the model from it. This method of curve fitting (by plotting the scatter diagram) which is done by eye is approximate only and depends upon the judgment of the operator and quite considerable variations can result from various individuals' efforts to draw the "best curve / straight line." This led to the development of the least square method of curve fitting and subsequently, software like MATLAB which help automate the process. This project work takes the automation process further by developing this software that seamlessly models a system and computes the level of agreement of the model developed using the data obtained from the field without any programming effort from the user.

Indexing terms/Keywords

Open Source, Modelling, Curve Fitting, Control.

Academic Discipline And Sub-Disciplines

Electrical and Electronic Engineering, Computer and Control Engineering

SUBJECT CLASSIFICATION

Modelling

Council for Innovative Research

INTRODUCTION

This work automates the modelling part of engineering and scientific tasks by developing this software that seamlessly models a system and computes the level of agreement of the model developed and the data obtained from the field without any programming effort from the user. This was achieved by working out the models for polynomial systems up to quartic level and exponential systems using the least square method of curve fitting. After these, the coefficient of determination, indicating the level of agreement of the model, was worked out and the resultant computation was used to develop a software using C# (C sharp programming language). Easy to use input and output screens were also developed and the software was tested using a manually worked out model and the result compared with the one computed using the software and results obtained were exactly the same.

In Engineering and Sciences, almost all the experimental data contain some error [1]. The purpose of employing the least square method of curve fitting was to devise a reliable method for establishing the relationship between two variables, corresponding values of which must have been obtained as a result of tests or experimentation [2]. These results in practice are highly likely to include some errors, owing to the imperfect materials used, the limitations of the measuring devices and the shortcomings of the operator conducting the tests and recording the result [3].

METHOD

A "model" or "empirical law" is a law which connects the two variables of a given data [4]. This project was achieved by working out the models for polynomial systems up to quartic level and exponential systems using the least square method of curve fitting. After these, the coefficient of determination, indicating the level of agreement of the model, was worked out and the resultant computation was used to develop a software using C# (C sharp programming language). Easy to use input and output screens were also developed and the software was tested using a manually worked out model and the result compared with the one computed using the software and results obtained were exactly the same.

The major forms of models considered are (knowing that they are the common forms encountered in practical scenarios):

$$y = ax + b \quad \text{(linear)} \quad (1)$$

$$y = ax^2 + bx + c \quad \text{(quadratic)} \quad (2)$$

$$y = ax^3 + bx^2 + cx + d \quad \text{(cubic)} \quad (3)$$

$$y = ax^4 + bx^3 + cx^2 + dx + e \quad \text{(quartic)} \quad (4)$$

$$y = a + be^{-x} + ce^{x} \quad \text{(exponential)} \quad (5)$$

Linear Systems

Considering a linear system of the form: $y = ax + b$

NB, for ingenuous reason, $\sum_{i}^{n}$ will be written as just $\sum$ for the rest of this work.

The regression line going by least square method of curve fitting is defined by;

$$\triangle = \begin{vmatrix} \Sigma x^2 & \Sigma x \\ \Sigma x & n \end{vmatrix} = n(\Sigma x^2) - (\Sigma x)^2$$

$$\triangle a = \begin{vmatrix} \Sigma xy & \Sigma x \\ \Sigma y & n \end{vmatrix} = n(\Sigma xy) - (\Sigma y)(\Sigma x)$$

Thus,

$$a = \frac{\triangle a}{\triangle} = a = \frac{n(\Sigma xy) - (\Sigma y)(\Sigma x)}{n\Sigma x^2 - (\Sigma x)^2} \quad (6)$$

and

$$\triangle b = \begin{vmatrix} \Sigma x^2 & \Sigma xy \\ \Sigma x & \Sigma y \end{vmatrix} = (\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy) \tag{7}$$

Thus,

$$b = \frac{\triangle b}{\triangle} = b = \frac{(\Sigma x)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2} \tag{8}$$

Quadratic Systems

For systems of the form, $\quad y \quad = \quad ax^2 + bx + c$

Resolving the above function by least square method of curve fitting yielded:

$$\Delta = \Sigma x^4 (n \Sigma x^2 - (\Sigma x)^2) - \Sigma x^3 (n \Sigma x^3 - (\Sigma x)(\Sigma x^2) + \Sigma x^2 ((\Sigma x^3)(\Sigma x) - (\Sigma x^2)^2) \tag{9}$$

$$\Delta a = \Sigma x^2 y (n \Sigma x^2 - (\Sigma x)^2) - \Sigma x^3 (n \Sigma xy - (\Sigma x)(\Sigma y)] + \Sigma x^2 [(\Sigma x)(\Sigma xy) - (\Sigma x^2)(\Sigma y)) \tag{10}$$
$$a = \Delta a/\Delta, \tag{11}$$

$$\Delta b = \Sigma x^4 [(n)(\Sigma xy) - (\Sigma x)(\Sigma y)] - \Sigma x^2 y(n (\Sigma x^3) - (\Sigma x)(\Sigma x^2)] + \Sigma x^2 [(\Sigma x^3)(\Sigma y) - (\Sigma xy)(\Sigma x^2)] \tag{12}$$
$$b = \Delta b/\Delta, \tag{13}$$

$$\Delta c = \Sigma x^4 [(\Sigma x^2)(\Sigma y) - (\Sigma xy)(\Sigma x)] - \Sigma x^3 [(\Sigma x^3)(\Sigma y) - (\Sigma x^2)(\Sigma xy)] + \Sigma x^2 y[(\Sigma x^3)(\Sigma x) - (\Sigma x^2)^2] \tag{14}$$
$$c = \Delta c/\Delta \tag{15}$$

Cubic Systems

These are systems of the form; $\quad y \quad = \quad ax^3 + bx^2 + cx + d$

Similarly,

$$\begin{aligned}
\Delta = \ & (\Sigma x^6)(\Sigma x^4)[n(\Sigma x^2) - (\Sigma x)^2] - (\Sigma x^6)(\Sigma x^3)[n(\Sigma x^3) - (\Sigma x)(\Sigma x^2)] \\
& +(\Sigma x^6)(\Sigma x^2)[(\Sigma x)(\Sigma x^3) - (\Sigma x^2)^2] - (\Sigma x^5)(\Sigma x^5)[n(\Sigma x^2) - (\Sigma x)^2] \\
& +(\Sigma x^5)(\Sigma x^3)[n(\Sigma x^4) - (\Sigma x)(\Sigma x^3) - (\Sigma x^5)(\Sigma x^2)[(\Sigma x)(\Sigma x^4) - (\Sigma x^2)(\Sigma x^3)] \\
& +(\Sigma x^4)(\Sigma x^5)[n(\Sigma x^3) - (\Sigma x)(\Sigma x^2)] - (\Sigma x^4)(\Sigma x^4)[n(\Sigma x^4) - (\Sigma x)(\Sigma x^3)] \\
& +(\Sigma x^4)(\Sigma x^2)[(\Sigma x^2)(\Sigma x^4) - (\Sigma x^3)^2] - (\Sigma x^3)(\Sigma x^5)[(\Sigma x)(\Sigma x^3) - (\Sigma x^2)^2] \\
& +(\Sigma x^3)(\Sigma x^4)[(\Sigma x)(\Sigma x^4) - (\Sigma x^2)(\Sigma x^3) - (\Sigma x^3)(\Sigma x^3)[(\Sigma x^2)(\Sigma x^4) - (\Sigma x^3)^2] \tag{16}
\end{aligned}$$

While,

$$\begin{aligned}
\Delta a = \ & (\Sigma x^3 y)(\Sigma x^4)[n(\Sigma x^2) - (\Sigma x)^2] - (\Sigma x^3 y)(\Sigma x^3)[n(\Sigma x^3) - (\Sigma x)(\Sigma x^2)] \\
& (\Sigma x^3 y)(\Sigma x^2)[(\Sigma x)(\Sigma x^3) - (\Sigma x^2)^2] - (\Sigma x^5)(\Sigma x^2 y)[n(\Sigma x^2) - (\Sigma x)^2] \\
& (\Sigma x^5)(\Sigma x^3)[n(\Sigma xy) - (\Sigma x)(\Sigma y)] - (\Sigma x^5)(\Sigma x^2)[(\Sigma x)(\Sigma xy) - (\Sigma x^2)(\Sigma y)]
\end{aligned}$$

$$(\Sigma x^4)\ (\Sigma x^2 y)\ [n(\Sigma x^3)\ -(\Sigma x)\ (\Sigma x^2)]\ -(\Sigma x^4)\ (\Sigma x^4)\ [n(\Sigma xy)\ -\ (\Sigma x)\ (\Sigma y)]$$

$$(\Sigma x^4)\ (\Sigma x^2)\ [(\Sigma xy)\ (\Sigma x^2)-\ (\Sigma x^3)(\ \Sigma y)-\ (\Sigma x^3)\ (\Sigma x^2 y)\ [(\Sigma x)(\Sigma x^3)-(\Sigma x^2)^2]$$

$$(\Sigma x^3)\ (\Sigma x^4)[\ (\Sigma x)\ (\Sigma xy)\ -\ (\Sigma x^2)\ (\Sigma y)]-(\Sigma x^3)\ (\Sigma x^3)\ [(\Sigma x^2)\ (\Sigma xy)-(\Sigma x^3)\ (\Sigma y)]\qquad(17)$$

$$a =\qquad \Delta a\,/\Delta \qquad\qquad(18)$$

Also,

$$\Delta b\ =\ (\Sigma x^6)\ (\Sigma x^2 y)\ [n\ (\Sigma x^2)\ -(\Sigma x)^2]\ -(\Sigma x^6)\ (\Sigma x^3)\ [n\ (\Sigma xy)-(\Sigma x)\ (\Sigma y)]$$

$$+(\Sigma x^6)\ (\Sigma x^2)\ [(\Sigma x)\ (\Sigma xy)-(\Sigma x^2)\ (\Sigma y)]-(\Sigma x^3 y)\ (\Sigma x^5)\ [n\ (\Sigma x^2)-\ (\Sigma x)^2]$$

$$+(\Sigma x^3 y)\ (\Sigma x^3)\ [n(\Sigma x^4)-(\Sigma x)(\Sigma x^3)]\ -(\Sigma x^3 y)\ (\Sigma x^2)\ [\ (\Sigma x^4)\ (\Sigma x)-\ (\Sigma x^2)(\ \Sigma x^3)]$$

$$+(\Sigma x^4)\ (\Sigma x^5)\ [n\ (\Sigma xy)-\ (\Sigma y)\ (\Sigma x)]\ -\ (\Sigma x^4)\ (\Sigma x^2 y)\ [n\ (\Sigma x^4)-(\Sigma x)\ (\Sigma x^3)]$$

$$+(\Sigma x^4)\ (\Sigma x^2)\ [\ (\Sigma x^4)(\Sigma y)-\ (\Sigma x^3)(\ \Sigma xy)]-(\Sigma x^3)\ (\Sigma x^5)\ [\ (\Sigma x)(\Sigma xy)-(\Sigma y)(\ \Sigma x^2)]$$

$$+(\Sigma x^3)\ (\Sigma x^2 y)[\ (\Sigma x)\ (\Sigma x^4)-\ (\Sigma x^3)\ (\Sigma x^2)]\ -(\Sigma x^3)\ (\Sigma x^3)\ [\ (\Sigma y)\ (\Sigma x^4)-\ (\Sigma xy)\ (\Sigma x^3)]\qquad(19)$$

$$b =\qquad \Delta b\,/\Delta \qquad\qquad(20)$$

And

$$\Delta c=(\Sigma x^6)\ (\Sigma x^4)\ [\ n\ (\Sigma xy)-(\Sigma x)\ (\Sigma y)]\ -\ (\Sigma x^6)\ (\Sigma x^2 y)\ [n\ (\Sigma x^3)\ -\ (\Sigma x)\ (\Sigma x^2)]$$

$$+(\Sigma x^6)\ (\Sigma x^2)\ [\ (\Sigma x^3)(\Sigma y)-\ (\Sigma x^2)\ (\Sigma xy)]\ -\ (\Sigma x^5)\ (\Sigma x^5)\ [n\ (\Sigma xy)-(\Sigma x)(\Sigma y)]$$

$$+(\Sigma x^5)\ (\Sigma x^2 y)\ [n\ (\ \Sigma x^4)-\ (\Sigma x)(\Sigma x^3)]\ -(\Sigma x^5)\ (\Sigma x^2)\ [\ (\Sigma x^4)\ (\Sigma y)-\ (\Sigma x^3)(\ \Sigma xy)]$$

$$+(\Sigma x^3 y)\ (\Sigma x^5)\ [n\ (\Sigma x^3)-(\Sigma x)\ (\Sigma x^2)\ -(\Sigma x^3 y)\ (\Sigma x^4)\ [n\ (\Sigma x^4)\ -\ (\Sigma x)\ (\Sigma x^3)]$$

$$+(\Sigma x^3 y)\ (\Sigma x^2)\ [(\Sigma x^4)\ (\Sigma x^2)\ -\ (\Sigma x^3)^2]\ -\ (\Sigma x^3)\ (\Sigma x^5)\ [\ n\ (\Sigma x^3)-(\Sigma x)(\ \Sigma x^2)]$$

$$+(\Sigma x^3)\ (\Sigma x^4)[n\ (\Sigma x^4)\ -\ (\Sigma x)\ (\Sigma x^3)]-(\Sigma x^3)\ (\Sigma x^2)\ [\ (\Sigma x^2)\ (\Sigma x^4)-\ (\Sigma x^3)^2]\qquad(21)$$

$$c =\qquad \Delta c\,/\Delta \qquad\qquad(22)$$

Also,

$$\Delta d=\ (\Sigma x^6)\ (\Sigma x^4)\ [(\Sigma x^2)\ (\Sigma y)\ -\ (\Sigma x)\ (\Sigma xy)]-(\Sigma x^6)(\Sigma x^3)\ [\ (\Sigma y)\ (\Sigma x^3)\ -\ (\Sigma x^2)\ (\Sigma xy)]$$

$$+(\Sigma x^6)\ (\Sigma x^2 y)\ [\ (\Sigma x)\ (\Sigma x^3)\ -\ (\Sigma x^2)^2]\ -\ (\Sigma x^5)\ (\Sigma x^5)\ [\ (\Sigma x^2)(\Sigma y)]\ -\ (\Sigma x)\ (\Sigma xy)]$$

$$+(\Sigma x^5)\ (\Sigma x^3)\ [\ (\Sigma x^4)(\Sigma y)-\ (\Sigma xy)(\Sigma x^3)]-\qquad\ (\Sigma x^5)\ (\Sigma x^2 y)\ [\ (\Sigma x)\ (\Sigma x^4)\ -\ (\Sigma x^2)(\ \Sigma x^3)]$$

$$+(\Sigma x^4)\ (\Sigma x^5)\ [\ (\Sigma y)\ (\Sigma x^3)\ -\ (\Sigma x^2)\ (\Sigma xy)]\ -\ (\Sigma x^4)\ (\Sigma x^4)\ [\ (\Sigma x^4)\ (\Sigma y)-\ (\Sigma x^3)(\Sigma xy)]$$

$$+(\Sigma x^4)\ (\Sigma x^2 y)\ [\ (\Sigma x^4)\ (\Sigma x^2)\ -\ (\Sigma x^3)^2]\ -\ (\Sigma x^3 y)\ (\Sigma x^5)\ [\ (\Sigma x^3)\ (\Sigma x)\ -\ (\Sigma x^2)^2]$$

$$+(\Sigma x^3 y)\ (\Sigma x^4)\ [\ (\Sigma x)\ (\Sigma x^4)-\ (\Sigma x^2)(\Sigma x^3)]\ -\ (\Sigma x^3 y)\ (\Sigma x^3)\ [\ (\Sigma x^4)\ (\Sigma x^2)\ -\ (\Sigma x^3)^2]\qquad(23)$$

$$d =\qquad \Delta d\,/\Delta \qquad\qquad(24)$$

Quartic Systems

These are system expressions of the form, $\quad y\ =\ a\,x^4+\ b\,x^3+\ c\,x^2\ +d\,x\ +\ e$

Similarly, the expression which defined the values of the constants (a, b, c, d and e) were carefully worked out and the resultant expressions obtained using least square method of curve fitting.

Exponential Functions

For exponential systems of the form; $\quad Y = a + be^{-x} + ce^{x}$

Applying the least-square curve fitting computation technique, it implies:

$$\Delta\ =\ n[(\Sigma e^{-2x})(\ \Sigma e^{2x}) - n^2] - (\Sigma e^{-x})[(\ \Sigma e^{-x})(\ \Sigma e^{2x}) - n(\Sigma e^{x})]$$

$$+ (\Sigma e^{x})[n(\Sigma e^{-x}) - (\Sigma e^{-2x})(\ \Sigma e^{x})]\qquad\qquad(25)$$

While,

$$\Delta a\ =\ \Sigma y[(\Sigma e^{-2x})(\ \Sigma e^{2x}) - n^2] - (\Sigma e^{-x})[(\ \Sigma ye^{-x})(\ \Sigma e^{2x}) - n(\Sigma ye^{x})]$$

$$+ (\Sigma e^{x})[n(\Sigma ye^{-x}) - (\Sigma e^{-2x})(\ \Sigma ye^{x})]\qquad\qquad(26)$$

$$a = \Delta a / \Delta \tag{27}$$

Also,

$$\Delta b = n[(\sum ye^{-x})(\sum e^{2x}) - n(\sum ye^{x})] - (\sum y)[(\sum e^{-x})(\sum e^{2x}) - n(\sum e^{x})]$$
$$+ (\sum e^{x})[(\sum e^{-x})(\sum ye^{x}) - (\sum ye^{-x})(\sum e^{x})] \tag{28}$$

$$b = \Delta b / \Delta \tag{29}$$

And

$$\Delta c = n[(\sum e^{-2x})(\sum ye^{x}) - n(\sum ye^{-x})] - (\sum e^{-x})[(\sum e^{-x})(\sum ye^{x}) - (\sum e^{x})(\sum ye^{-x})]$$
$$+ (\sum y)[n(\sum e^{-x}) - (\sum e^{-2x})(\sum e^{x})] \tag{30}$$

$$c = \Delta c / \Delta \tag{31}$$

Coefficient of Determination

This measure denotes the level of agreement of the model. It calculates what proportion of the variation in the actual value of y may be predicted by changes in the value of x. It is a square of the correlation coefficient.

It is defined by

$$\frac{Explained\ variation}{Total\ variation} = \frac{\sum (YE - \ddot{y})}{\sum (y - \ddot{y})} \tag{32}$$

Where,

YE = estimate of y (derived by feeding x into the model)

y = individual value of y

$\ddot{y}$ = mean of actual value of y

SYSTEM DESIGN

The corresponding program codes for the various functions derived earlier (see method, above) were then developed along with simple user input and output interfaces in C# programming language.
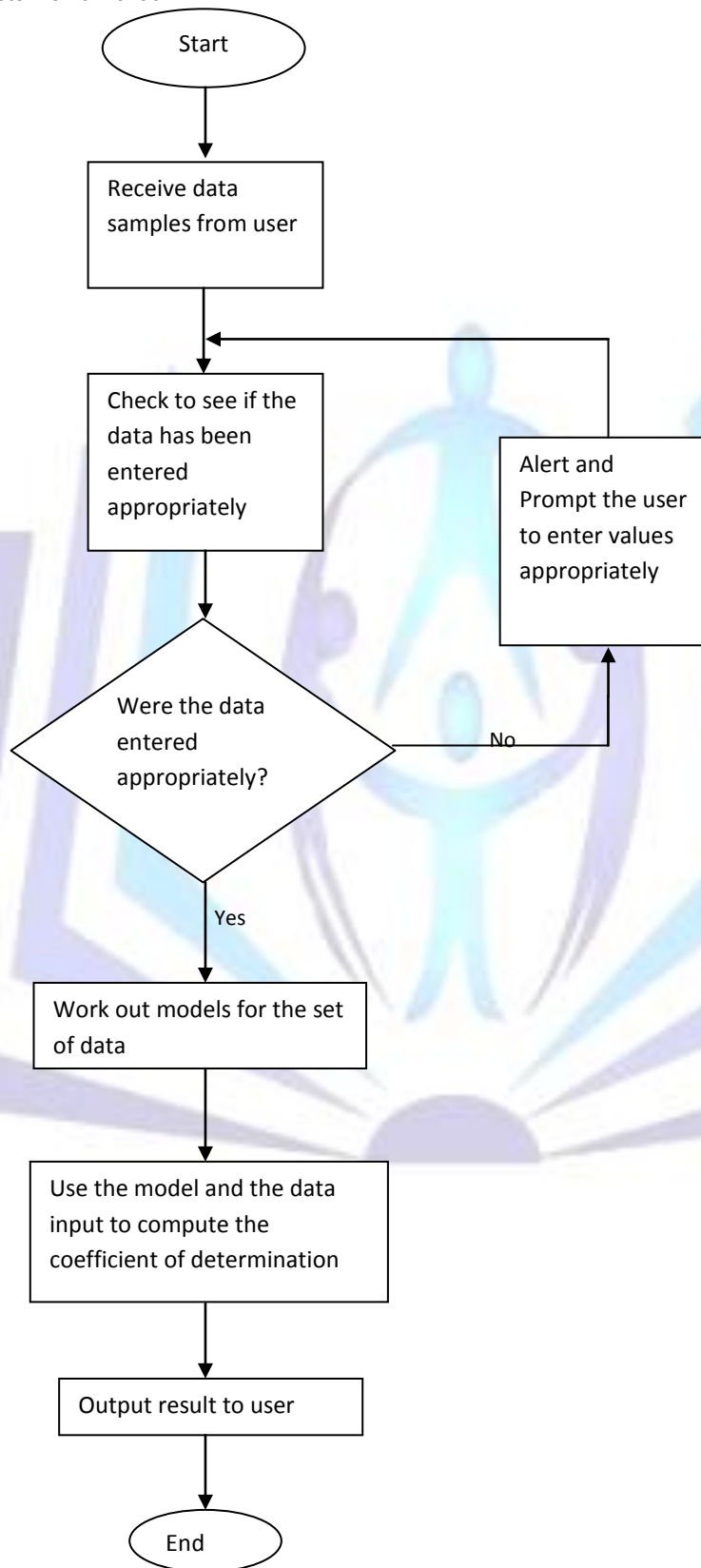
Below is the system's flow chart.

Fig. 3.1: Overall program flow-chat

When the program is started, it follows the pattern illustrated in the flow-chat shown in fig. 3.1 to work out the model of the given data.

It starts off by presenting the user with an interface (input form) to enter the data obtained from experimentation. It then proceeds to check if the user entered the data appropriately by running a consistency test, if the data was entered appropriately, the software proceeds automatically to work out the model for each of the systems defined earlier and also the level of agreement the model has with the experimental result (that is, the coefficient of determination) and presents both the model and the coefficient of determination (the level of agreement expressed in percentage) in the output form. If on the other hand, the data entered is found to be inconsistent, the software prompts the user to enter the data appropriately by indicating the field requiring proper data entry before proceeding to work out the model and computing the coefficient of determination.

VALIDATION OF THE SOFTWARE USING OHM'S LAW

The simple circuit, shown below, was set up in ELECTRONIC WORKBENCH.
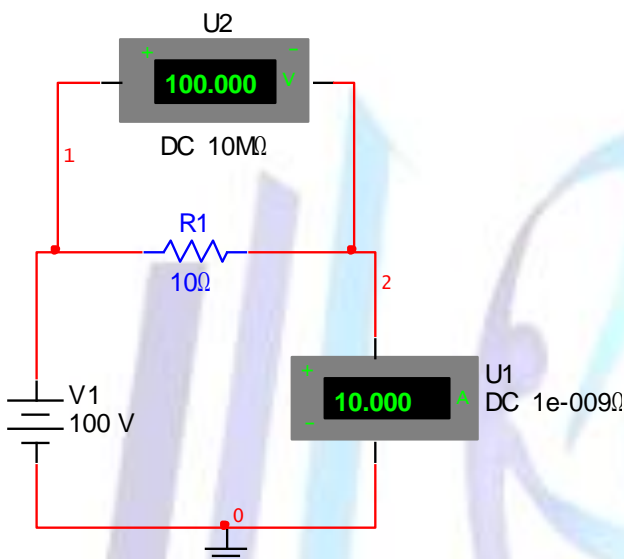

Figure 4.1: Circuit setup in ELECTRONIC WORKBENCH

With a fixed resistor of resistance 10 ohms, different supply voltages, V, where used to simulate the circuit and the corresponding current, A, as read by the ammeter were recorded and tabulated as shown below:

Table 4.1: Table of values for V and I using a fixed 10ohms resistor

V (in volts)	100	75	50	25	10
I (In Amps)	10.0	7.5	5.0	2.5	1.0

The different values shown above, Table 4.1, were fed as input to the modelling software via the input form and run. The program gave;

$V = 10I$ (33)

Equation (33) if observed closely is exactly ohms law (V=IR; mathematically) where R = 10

SUMMARY OF TEST RESULTS

The software was tested with the values in *table 5.1* and the results were exactly the same with the one obtained in the analysis above.

Table 5.1: table of values of samples

X	10.0	7.5	5.0	2.5	1.0
Y	100	75	50	25	10

The corresponding polynomial system result from the software was;

$$Y = 10X \qquad\qquad (34)$$

with a coefficient of determinationof (percentage of agreement)100%.

This was exactly the same as the result obtained from the manual analysis.

CONCLUSION AND RECOMMENDATION

Conclusion

The task of establishing a relationship between system variables (especially non-linear system) could be cumbersome, very technical, time consuming and error prone. Thus, there is a need to automate the modelling process.

This modelling software has been developed to address the aforementioned short comings.

Recommendations

This software is strongly recommended for use in academics (universities, polytechnics, research institutes,…) for teaching staff, students and research officers alike. It is also recommended for use by engineers and scientists in real life / field practice.

It is also recommended that during the course of data / sample collection, it should be collected over the operating range (the operating limits of the equipment / study should be included) if possible.

It is also recommended that adequate number of samples should be collected. The more the number of samples collected, the better the model.

It is advised to always select the model that gives the best coefficient of determination because it provides the reliable model.

It is a good practice to only select a model if its coefficient of determination is above 75%.

REFERENCES

[1]   Ogbonnaya I. Okoro, "Introduction to MATLAB/SIMULINK for Engineers and Scientists," John Jacob's Classic Publishers Ltd. Enugu, 2005, pp. 85 – 88.

[2]   K.A. Stroud, and Dexter J. Booth, "Engineering Mathematics," PALGRAVE Houndmills New York, 2001, pp. 709 – 718.

[3]   K.A. Stroud, and Dexter J. Booth, "Engineering Mathematics," PALGRAVE MACMILLAN Houndmills New York, 2007, pp. 719 – 740.

[4]   H.K. Dass, "Advanced Engineering Mathematics," S. Chand and Company New-Delhi, 2010, pp.1232-1242.