EPH - International Journal of Applied Science
ISSN (Online): 2208-2182
Volume 01 Issue 04-December, 2025

DOI: https://doi.org/10.53555/eijas.v1i4.19

FORECASTING OF THE MAXIMUM ANNUAL FLOOD HEIGHT OF A RIVER USING THE LEAST SQUARES METHOD

K.Binyahmed[1*], H. Said [2]
*[*1] Mathematics, Science/University Sebha, Libya*
[2] Electrical Engineering, Higher Institute of Technology University Sebha, Libya

[*]Corresponding author:-
***E-mail**: kham243@yahoo.com*

Abstract:-
This paper deals exclusively with mathematical functions that link between time and the maximum annual flood height of Ocmulgee River in the U.S. state of Georgia during 60 years.
The best equations used this study were the cftool program including with the Matlab program. This program is based on the least square method using ready functions or suggesting any other functions. We looked at the data as a whole and found the best curve with the smallest errors. The second method was to look at the data horizontally. The differences between the years are 10 years, because there is a close approximation of these data. The results found that the error in the second method was much less than the error in the first method and the error did not exceed 1.6%.

Keywords: - *Curve fitting, Least Square method.*

INTRODUCTION

In different experiments and research observations , we often obtain a set of simultaneous readings of two or more varable It may then be useful to find the relationship between the variables in addition that correspond to these corresponding values in a function that links the change that occurs as a result of other changes[1-4].

And this is translated mathematically by the following relationship:

$$y = f(x_1, x_2, x_3, \dots) \qquad (1)$$

where as $(x_1, x_2, x_3, \dots)$ are independent variables, y is dependent variable and f Is the function that links the variables. The simplest form of the previous relationship is using one independent variable in addition i.e., the relationship between two variables is only the form of [4, 5]

$$y = f(x) \qquad (2)$$

In practical experiments, we obtained a table for a set of corresponding values of the independent variable x and the dependent variable y

in addition each of these symmetrical values represents a point on or near a curve [4, 5]. The next important thing is to obtain the equation of the curve that passes through all or most of these points or close to them so that this equation shows the general picture of the relationship between the two variables of the previous relation studied, and the equation we obtained in this way is called Empirical Equation for the curve in addition this method is called: (Curve Fitting) [2, 6].

We may find that there are more than one image blobs group compatibility that could draw a straight line or curves, each of which corresponds to the set of points appropriately in addition therefore, from the beginning, the shape of the curve that one expects to obtain must be chosen based on the study and knowledge of the theoretical considerations of the relationship between the variables and their natural meaning [2, 6].

In general, if the function is not known, more than one equation can be reconciled to the same experimental results, and it is preferred to be equivalent to the other by the accuracy of its representation of the results (correlation coefficient) or the simplicity of its use[2, 6, 7].

The most important equations are the following:

1- Linear Equation $y = a_0 + a_1 x$ (3)

2- Parabolic Equation $y = a_0 + a_1 x + a_2 x^2$ (4)

3 - Exponential Function

$y = a_0 e^{a1x}$ (5)

4- Power Function

$y = a_0 x^{a1}$ (6)

5 - Poisson Function

$y = a_0 x e^{-a1x}$ (7)

6- Gauss Function

$y = a_0 x e^{-a1x2}$ (8)

The steps that follow the selection of the curve function are finding the values of the constants in one of the following ways: 1- Graphical method

2- Average Points method

3 - Least Square method

The graphical method is based on personal judgment in drawing an approximate curve to reconcile a set of data, which is called the hand-to-hand method of curve adjustment. While the intermediate point method is simple and easy to apply in addition the least square method gives a result more accurately than the previous two methods [4, 6, 8].

Exponential Function

$y = a_0 e^{a1x}$ (5)

We can evaluate this equation (converted it into a straight-line equation) by taking logarithms for two sides of the equation. The exponential function is the reverse of the logarithmic function in addition

$\ln y = \ln a_0 + a_1 x$ (9)

This is the equation of the intersection of line with the vertical axis

$\ln a_0$ and tilt line a_1 in addition a_0 and a_1 can usually be determined from the straight line drawing [4, 6, 8].

Method of Least squares

Definition of deviations

Assume that it is in Figure (1), A curve $y = f(x)$ the data points

$P1\ (x1, y1), P2(x2, y2), P3(x3, y3)$

And defining deviation for any of the points shown in the form of the appropriate curve as the vertical distance.

Between this point and curve [1].

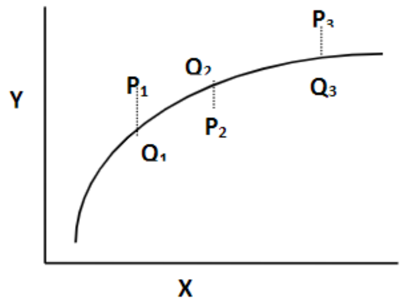


Figure (1) a curve between this points

So the deviation of the three points P_1 , P_2 ,P_3 respectively is :

$Q1P1 = y1 - f(x1)$

$Q_2P_2 = y_2 - f(x_2)$ (10)

$Q3P3 = y3 - f(x3)$

The deviation is positive if the point is above the curve and negative if the point is below the curve. From this definition we see that the best fit for the curve is where the absolute values of deviations are small in addition

It can be seen that the straight line that corresponds to the readings given in the method of the intermediate points is what divided the readings into two categories is the straight line in which the algebraic sum of the deviations is equal to zero [4, 6, 8]. This does not mean that in the way of the intermediate points, it is necessary to obtain the best fit. The absolute values of the deviations may be large, although the positive deviations are the balance of the negative deviations. Therefore, the algebraic sum of deviations cannot be used as a principle of good curve [4].

Since squares of deviations, however, are always positive, then the sum of squares of deviations is not equal to zero (unless the curve passes through all points) in addition thus, it is clear that the best fit is when this sum (total squares of deviations) is small) in addition Thus, the curve in a given form for a class of points is determined by the constants so that the sum of the squares of deviations is small whenever possible [4, 6].

Method of Least Squares for Linear Type*:*

Assume that the form of the linear equation is:

y = $a_0 + a_1$x (3) If the readings given are (x_i, y_i) $(i = 1,2,3,4, \dots, n)$ as defined by deviations, the sum of squares of deviations is equal to:

$$E = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \qquad (11)$$

It is clear that E is a function in a_0 and a_1). To find the minimum value the partial differential coefficients relative to both a_0 , a_1 will be equal to zero. That is:

$$\frac{\partial E}{\partial a_0} = 0 , \frac{\partial E}{\partial a_1} = 0 \qquad (12)$$

$$or \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i) , \sum_{i=1}^{n} x_i (y_i - a_0 - a_1 x_i) \qquad (13)$$

The last two equations can be written as follows:

$$\sum_{i=1}^{n} x_i y_i = a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 \qquad (14)$$

$$\sum_{i=1}^{n} y_i = a_0 n + a_1 \sum_{i=1}^{n} x_i \qquad (15)$$

Where n is the Number of readings given in addition these equations are called Normal equivalents. And by solving these equations, we get the values of a_0 , a_1 [4, 6, 8] .

The least squares method for a polynomial of degree m:

If we have the readings:

(x_i, y_i) $(i = 1, 2,3,4, \dots, n)$, the curve of these readings is to be reconciled to many polynomial m where $m < (n - 1)$ that means many limits on the form:

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = \sum_{k=0}^{m} a_k x^k \qquad (16)$$

In this case, the sum of squares of deviations is equal to:

$$E = \sum_{i=1}^{n} \left(y_i - \sum_{k=0}^{m} a_k x_i^k \right)^2 \qquad (17)$$

It is clear that **E** Function in a_k where $k = 0, 1, 2, 3, \dots, m$ In order **E** is the least that can be:

$$\frac{\partial E}{\partial a_k} = 0 \quad , (k = 0,1,2,3, \dots, m) \qquad (18)$$

And these give us (m + 1) of the equations and their solution gives us values of the constants [5, 6, 8].

$$a_k \quad , (k = 0,1,2,3, \dots, m)$$

Materials and methods of work:

The analysis of the data was done by the use of (cftool) program included in Matlab (version 13). The data the maximum values of the annual flow of a river for 60 years were taken from [9].

Table (1) Yearly Maxima Wind Data

51	41	31	21	11	1
32.88	31.60	30.32	29.48	26.46	24.21
35.89	35.59	35.23	35.14	33.63	33.03
37.48	37.13	36.50	36.49	36.07	35.95
39.45	39.26	38.91	38.53	38.21	38.01
41.05	41.03	40.69	40.49	40.36	40.32
43.31	43.05	42.91	42.82	42.62	41.54
45.04	44.71	43.87	43.65	43.42	43.34
49.43	49.28	48.76	48.29	46.00	45.58
52.94	52.54	51.90	50.73	50.45	50.17
75.70	67.76	61.95	60.10	57.84	54.01

Figure (2) represents the data in Table (1) the curve is periodically and has a 10 - year cycle period, but differs in amplitude so that the amplitude increases each cycle. Notes the similarity function tan (x).

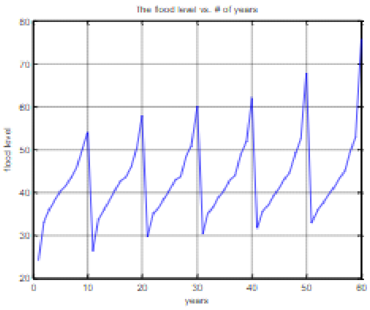
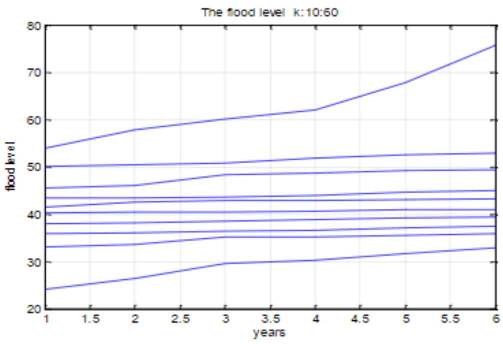


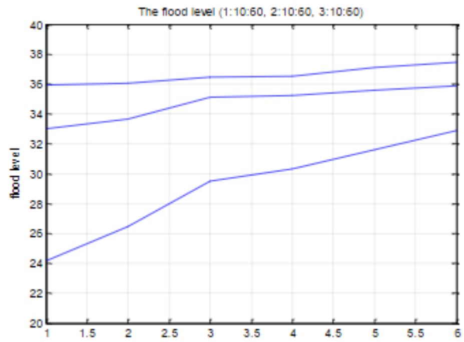
Figure (2) years and flood level

Results and discussion:

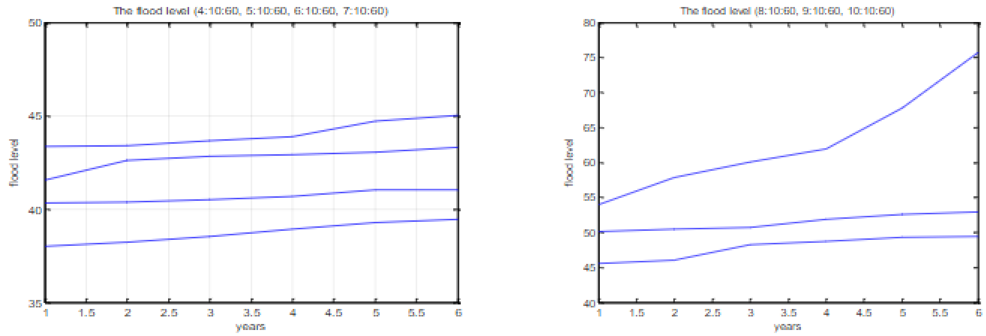
Figure (3) (a,b,c,d) Shows the data when we toke the horizontal years as 11, 21,31,41,51. Each curve can be presented by certain equation. All curves look like smooth curves.



a



b



c

Figure (3) (a,b,c,d)

The results:

Using (cftool) for each curve in the two cases:

The first cases of looking at the data as in Figure (2)

(1) the data have been normalized to deal with small values graphically. The error maximum reaches 25%.

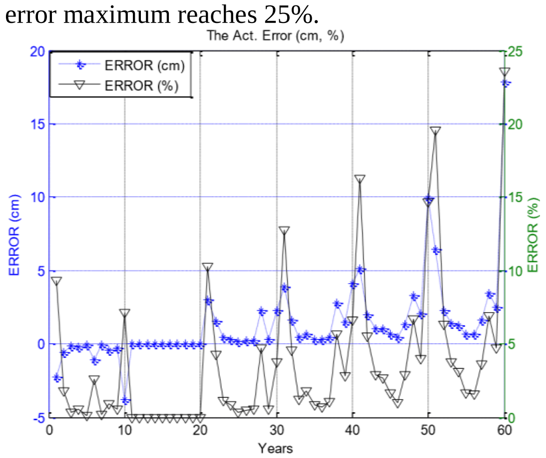


Figure (4) years and Error

(B) The data were then processed using (cftool).The error has been reduced to about (5 %) as shown in Figure (5)

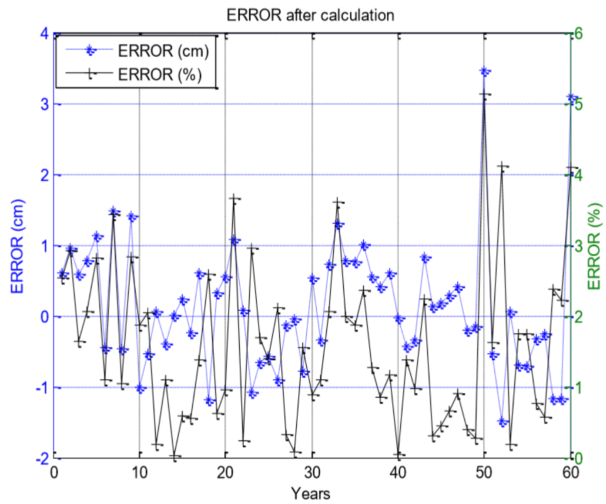


Figure (8) years and Error

The second method **is to** use horizontal data, where the difference between a point and the next is10 years. By looking for the appropriate equations, the values of the errors are the least. This work had been divided into stages:

(a) Use data for the first 40 years, then predict the last 20 years of (41:60). From the data in figure (6) the maximum od error was less than (2 %).

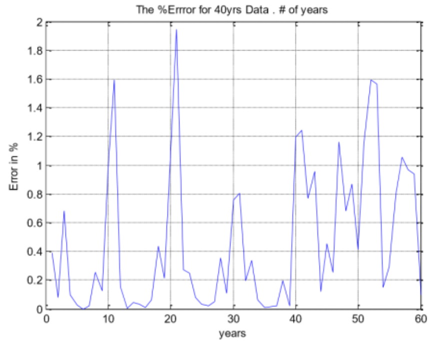


Figure (6) years and Error

(B) Use the data for the first 50 years, and then predict the 10 years of the last years of (5 1:60). We have noticed that the maximum of errors is less than (1.8 %) as shown in Figure (7).

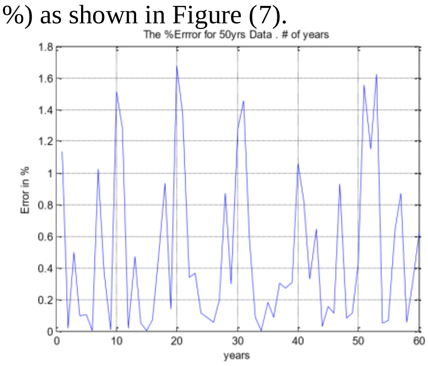


Figure (7) years and Error

(C) Using the data as a whole for 60 years. In Fig. (8) , the errors are significant at the end of the periods, and the processing was done using exponential functions, such as the maximum of error was less than (1.6%)

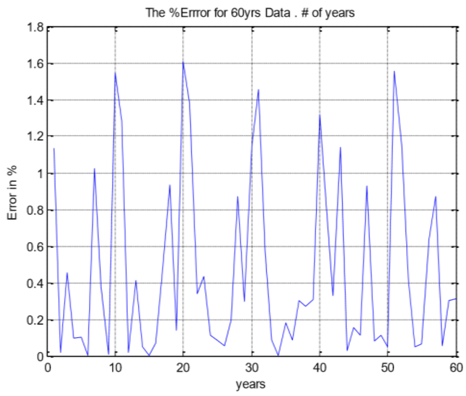


Figure (11) years and Error

References:

[1].**McClarren, R., (2018).** Computational Nuclear Engineering and Radiological Science Using Python, Academic Press.

[2].**Molugaram, K., et al., (2017).** Statistical techniques for transportation engineering, Butterworth-Heinemann.

[3].**Johnson, M. L., (2008).** Nonlinear Least-Squares Fitting Methods. Methods in Cell Biology **84**: 781-805.

[4].**Wylie, C. R., (1960).** Advanced engineering mathematics.

[5].**Stark, P. A., (1970).** Introduction to numerical methods.

[6].**Sastry, S. S., (2012).** Introductory methods of numerical analysis, PHI Learning Pvt. Ltd.

[7].**Sekiya, F. and A. Sugimoto., (2012).** Discrete Polynomial Curve Fitting to Noisy Data. International Workshop on Combinatorial Image Analysis, Springer.

[8].**Richard, L. and J. Burden.,(2000).** Douglas Faires. Numerical ANALYSIS [M]. Thomson Learning. Inc. Aug.

[9].**Castillo, E., et al., (2005).** Extreme value and related models with applications in engineering and science.