MATLAB-Based Standalone Application for Estimation of Growth and Mortality Parameters of Fish

Ivelina Zlateva[1], Nikola Nikolov[2], Mariela Alexandrova[3]

Department of Automation, Technical University of Varna, Bulgaria

Abstract— *The exploited fish populations and the fishery dependent on them require regular monitoring and periodical stock assessment. The resource evaluation is crucial for definition of appropriate management strategies for rational exploitation and long-term sustainable development of the exploited stocks.*

The determination of the current state of a given stock is directly related to analysis of historical and present data concerning specific stock parameters which indicates the current status of the stock and suggests relevant definition of healthy exploitation levels. Such key stock parameters are growth and mortality rates, abundance indices, carrying capacity, maturity and reproduction, catch composition and recruitment [1,3,4,6,13,14].

The present article introduces a program system - MATLAB-based standalone application specifically developed to facilitate length-weight relationship of fish analysis outcome, growth and mortality parameters estimates and their confidence limits, normality tests of length-weight-frequency sample data, used in analysis. The proposed system incorporates statistical analysis of the overall significance of the mathematical models and results delivered, as well as methodology to overcome added computational errors in numerical analysis.

Keywords— *fish stocks, growth and mortality parameters, MATLAB programming environment, standalone application, stock assessment.*

I.   INTRODUCTION

Background

The present material introduces a MATLAB-based standalone application specifically developed to support the fish stock assessment procedure. It can be executed in MATLAB environment as well by simply typing the name of the program GR_MORT_PAR_EST in MATLAB workspace. It is focused on length-weight relationship analysis, von Bertalanffy individual growth parameters and mortality parameters estimation, based on length-frequency, length-weight frequency, length-at-age and catch-at-age samples data, by applying variety of methods introduced in fisheries science, which can be selected depending on the input data type and availability. Another incorporated feature is related to validation of the accuracy of the estimates provided by using detailed statistical analysis of the adequacy and the overall significance of the mathematical models obtained, construction of confidence intervals of the parameters estimated and validation of specific distributional assumptions of the sample data under analysis.

The below listed methods can be selected for execution in accordance with the data collected [4, 13]:

II.   METHODS

2.1   Length-weight relationship analysis

2.1.1   Investigation of linear length-weight relationship model validity using linear regression analysis of L(x) as independent variable and W(y) – as a dependent variable (direct measurements)

Input Data: Length (cm) and Weight (g) measurements (length-weight frequency samples)

Deliverables: Proves or rejects validity of a linear length-weight relationship model of a kind: $W(i) = a + b * L(i)$, where: W is weight of fish, L is the length of fish and constructs confidence limits for the regression coefficients a and b. Provides graphical representation of the model fit over the actual (measured) data.

2.1.2 Investigation of non-linear length-weight relationship model validity eq. (1.0) by linear transformation of (1.0) an linear regression of log transformed length-weight data, analyzing L(x) as independent variable and W(y) – as a dependent variable

Input Data: Length (cm) and Weight (g) measurements (length-weight frequency samples)

Deliverables: Proves or rejects the validity of a model of kind: $W(i) = q * L(i)^b$, Where: $W(i)$ is the body weight of a fish, $L(i)$ is the total length and q and b are parameters, provides estimates of and q and b, once the validity of the model is justified and constructs confidence limits for the coefficient b. Provides graphical representation of the model fit over the actual (measured) data.

2.2 Estimation of growth parameters by using the Von Bertalanffy equation, which expresses the length, L as a function of the age t of the individual:

$$L(t) = L_\infty(1 - exp(-K(t - t_0)),$$

Where: L_∞ is the asymptotic length, K is a parameter, which describes how fast the individuals approaching L_∞ and t_0 is the initial condition parameter, or determines the time when the individual has zero length.

Input data: age readings (years) and Length measurements (cm)

Deliverables: Provides estimation of K (coefficient which shows how fast the species investigated approaches the asymptotic (infinite) length L_∞), estimates of the asymptotic length L_∞ and constructs confidence limits for the coefficient K. Provides graphical representation of the model built with the parameters estimated.

Methods for estimation of growth parameters from length at age data:

2.3 Estimation of K and L$_\infty$ with the Gulland Holt Plot

Input data: age readings (years) and Length measurements (cm)

Deliverables: Provides estimation of K (coefficient which shows how fast the species investigated approaches the asymptotic (infinite) length L_∞) , estimation of asymptotic length L_∞ and constructs confidence limits for the coefficient K. Provides graphical representation of the model fit over the actual (measured) data.

2.4 The Ford-Wallford Plot

Input data: age readings (years) and Length measurements (cm)

Deliverables: Provides estimation of K (coefficient which shows how fast the species investigated approaches the asymptotic (infinite) length L_∞) , estimation of asymptotic length L_∞. Provides graphical representation of the model fit over the actual (measured) data.

2.5 The Chapman's method

Input data: age readings (years) and Length measurements (cm)

Deliverables: Provides estimation of K (coefficient which shows how fast the species investigated approaches the asymptotic (infinite) length L_∞) , estimation of asymptotic length L_∞. Provides graphical representation of the model fit over the actual (measured) data.

2.6 Estimation of K and t_0 with the von Bertalanffy plot

Input data: age readings (years), Length measurements (cm) and L_∞ - the asymptotic length must be a priori known

Deliverables: Provides estimation of K (coefficient which shows how fast the species investigated approaches the asymptotic (infinite) length L_∞) and t_0 (the *zero* age for the species under investigation)

Estimation of Mortality Rates

2.7   Estimation of Z (total mortality rate) by using the Linearized Catch Curve based on age composition data

Input data: age groups t (years), numbers caught per year per age groups $C_{(y,t,t+1)}$

Deliverables: Provides an estimate of Z (total mortality rate) and constructs Z confidence limits

2.8   Estimation of Z (total mortality rate) by using the Linearized Catch Curve based on length composition data

Input data: Length groups $(L_1 - L_2)$, catch in numbers per length group $C_{(L_1-L_2)}$, K and L_∞;

Deliverables: Provides an estimate of Z (total mortality rate) and constructs the confidence limits

2.9   Beverton and Holt Z-equation based on age composition data

Input data: Length groups $(L_1 - L_2)$ above chosen length L', catch in numbers per length group $C_{(L_1-L_2)}$, K, L_∞ and L';

Deliverables: Provides an estimate of Z (total mortality rate) and constructs Z confidence limits

2.10   The cumulated curve based on length composition data – The Jones Van Zalinge method

Input data: the lower value of Length groups $(L_1 - L_2)$ intervals L_1, the cumulated catch in numbers per length group $C_{(L_1,L_\infty)}$, K and L_∞

Deliverables: Provides an estimate of Z (total mortality rate)and constructs Z confidence limits

2.11   The Powell-Wetherall Method

Input data: the lower value of Length groups $(L_1 - L_2)$ intervals L_1, and $\bar{L}$ - the mean length of a fish longer than L';

Deliverables: Provides an estimate of Z/K, L_∞ and constructs confidence limits for the regression coefficients obtained

2.12   Estimation of M and q by using fishing effort and Z estimates

Input data: the fishing effort$f(y)$, L_c - the 50% retention length and $\bar{L}_c$ - the mean length of all individuals presented in the catch;

Deliverables: Provides an estimate of M(natural mortality rate) and q - the catchability coefficient, constructs confidence limits for M and q

2.13   Normality test of length frequency samples

Input data: Length (weight) measurements (cm);

Deliverables: the method execution render automatically a statement, based on the χ^2 normality test result for retention or rejection of the null hypothesis H_0, that the sample data follows the normal (Gaussian) distribution as automatic building of histograms and probability plots.

III.   CALCULATION PROCEDURES AND ALGORITHMS USED IN THE PROGRAM SYSTEM DEVELOPMENT

Most of the methods described above are based on parameters estimation by applying linear regression analysis (methods 1 – 12) [1,4,6,13,14]. Naturally the models analyzed are best described with non-linear relationships, which are further proved linear in the model coefficients on conditions and after certain linear transformations, hence easily estimable by using regression analysis. Furthermore, regression analysis estimates the parameters of the regression equation that best describes the relationship between two variables. In this case, "best" is defined in terms of the linear regression equation that minimizes the sum of squared errors of prediction and, therefore, maximizes the coefficient of determination. The program system is developed in a manner not only to render automatically parameter estimates and mathematical models, but to investigate further the overall model significance and its validity in terms of determination and adequacy of the models obtained by following the next steps:

3.1   Determine regression coefficients of the model:

$$\hat{y}_j = \sum_{i=0}^{k} b_i * f_{ji} + \varepsilon_j \tag{1}$$

➢ **Form the input-output data vectors:**

$$x = [x_1, x_2, x_3, \dots x_n]^T \text{ - input vector of the independent variable } x \tag{2}$$

$$y = [y_1, y_2, y_3, \dots y_N]^T \text{ - output vector of the dependent variable } y \tag{3}$$

➢ **Form the experiment matrix X and the regression matrix F:**

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{23} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{Nn} \end{bmatrix}; F = \begin{bmatrix} f_{10} & f_{11} & \dots & f_{1k} \\ f_{20} & f_{21} & \dots & f_{2k} \\ \dots & \dots & \dots & \dots \\ f_{N0} & f_{N1} & \dots & f_{Nk} \end{bmatrix} \tag{4}$$

➢ **Determine the coefficients bᵢ of the model:**

$$b = C * F^T * y \tag{5}$$

where: $C = (F^T * F)^{-1}$ is the covariance matrix, on condition that the determinant of $(F^T * F)$ must be $\neq 0$ and its condition number $< 10\text{-}10^2$ [2,5,11].

If matrix $(F^T * F)$ is identified ill-conditioned during the execution of the program it will automatically standardize (or normalize) the variables used in regression analysis: $x_i^{norm} = \frac{\tilde{x}_i - \bar{x}_i}{\sigma_{\tilde{x}_i}}$, where: $\tilde{x}_i$ is the raw (measured) value of the input variable, $\bar{x}_i$ is the mean and $\sigma_{\tilde{x}_i}$ is the standard deviation [7]. The analysis with implemented normalized variables prove to facilitate condition number of the matrix$(F^T * F)_{norm} < 10 - 10^2$, which in turn leads to conclusion that the algorithm delivers an approximation of the solution whose precision is no worse than that of the data or it does not introduces errors of its own [5,11].

3.2   Analysis of the overall statistical significance of the mathematical model delivered:

➢ **Calculate the total sum of squares Q:**

$$Q = \sum_{i=1}^{N} (y_i - \bar{y})^2 \tag{6}$$

Where $\bar{y}$ is the mean value of the output variable and is obtained with: $\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$

➢ **Calculate the partition of the sum of squares:**

$$Q_{part} = \sum_{i=1}^{N} (y_i - \hat{y})^2 \tag{7}$$

Where: $\hat{y}$ are the values of the output variable assessed by the model delivered.

➢ **Calculate the correlation coefficient R, which is a measure for determination of the mathematical model obtained;**

$$R = \sqrt{1 - \frac{Q_{part}}{Q}} \tag{8}$$

➢ **F-test of the overall model significance:**

$$F = \frac{R^2(N-k-1)}{(1-R^2)k} \tag{9}$$

The critical value $F_c(\alpha, \vartheta_1, \vartheta_2)$ is found in F-distribution table for significance level $\alpha = 0,05$ and degrees of freedom $\vartheta_1 = k$ and $\vartheta_2 = N - k - 1$.

If $F > F_c$, the calculated value of R is assumed significant for $\alpha = 0,05$, and proves that the model delivered by using the above described calculation procedure is adequate [7].

The last method of the above listed (13) - Normality test of length frequency samples is developed to justify the assumption that the frequencies in length-frequency samples used for analysis in fish stock assessment procedure follow approximately the normal distribution by using the chi-square normality or goodness of fit test.

The probability density function of the normal distribution is:

$$f(x|M,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} exp[-\frac{(x-M)^2}{2\sigma^2}] \tag{10}$$

where: the expectation M (also median and mode) and the standard deviation σ are distribution parameters, which characterize the centre of the distribution and its scale and σ^2 is the variance of M:

$$M = \int_{-\infty}^{\infty} x.f(x)dx \tag{11}$$

$$\sigma^2 = \int_{-\infty}^{\infty}(x-M)^2.f(x) \tag{12}$$

here: $-\infty < x > \infty,\ -\infty < M > \infty,\ \sigma > 0.$

Significant and unbiased estimates of the expectation M and variance σ^2 when the sample is broken to k-intervals (where: $k \approx 1 + 3.22 * log_{10}(n)$ and n is the number of observations) are:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i^* n_i = \sum_{i=1}^{k} x_i^*.P_i \tag{13}$$

where: $P_i = \frac{n_i}{n}$ is the empirical probability distribution function and the calculated variance:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i.(x_i^* - \bar{x})^2 = \frac{n}{n-1}\sum_{i=1}^{k}(x_i^* - \bar{x})^2.P_i \tag{14}$$

Where: x_i^* is the mid-point of the „ i-th" interval, and n_i are the observed frequencies in a given interval.

Once the data is processed and the expectation and observed frequencies are calculated as per the above listed mathematical procedure, the theoretical (expected) frequencies n_{i_t} are obtained by solving:

$$f(x) = \frac{n*dl}{s\sqrt{2\pi}} exp[-\frac{(x-M)^2}{2s^2}], \text{ where } dl \text{ is the interval size.}$$

The experimental data is tested further for normality using the chi-square normality test which is being used to test if a sample of data came from a population with a specific distribution [9,10,12]:

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n_{i\_t})^2}{n_{i\_t}} \tag{15}$$

where: n_{i_t} are the theoretical frequencies, n_i are the observed frequencies [8].

If the calculated χ^2 value is below a critical (table) value $\chi_T^2(v;p)$ for significance level α, (where v are the degrees of freedom and p is the probability) the null hypothesis is retained, stating that the sample data is taken from a population with specific distribution (normal distribution).

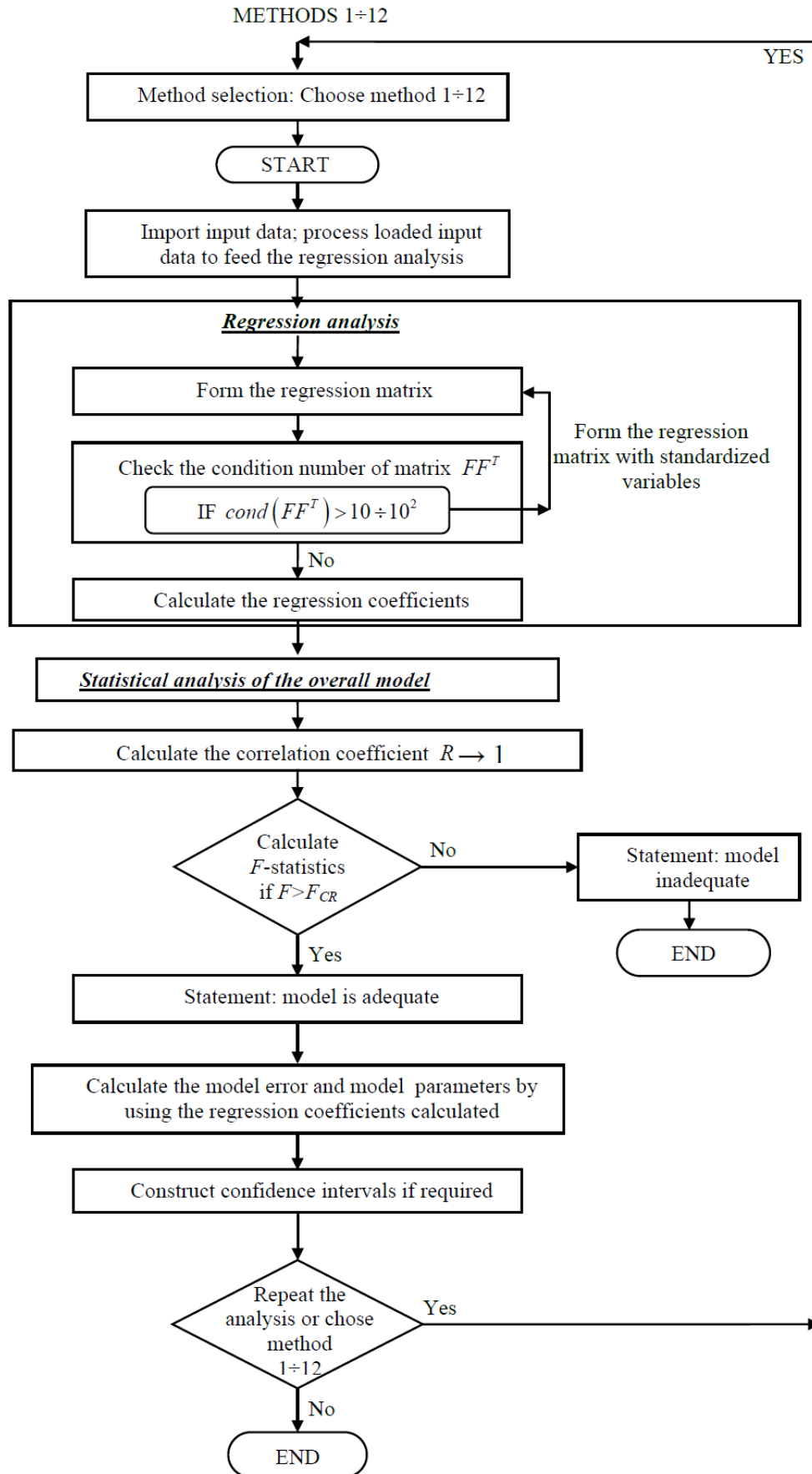Algorithm block-diagrams are presented in Fig. 1 for methods 1 to 12 and in Fig. 2 for method 13.

METHODS 1÷12

FIGURE 1. ALGORITHM BLOCK-DIAGRAM – METHODS 1 TO 12

METHOD 13

Method 13 selected for execution:

START

import input data

Calculate the number of bins

Build the histogram and calculate the observed frequencies n_i

Calculate the empirical probability, expectation and variance

Calculate the theoretical (expected) frequencies n_{i_t}

Combine intervals with $n_{i\_t} < 5$

Calculate χ^2 value

$\chi^2 < \chi_T^2(v, p)$

YES → Statement: the Sample data follow specific distribution (normal distribution)

NO

Statement: the Sample data do not follow the normal distribution
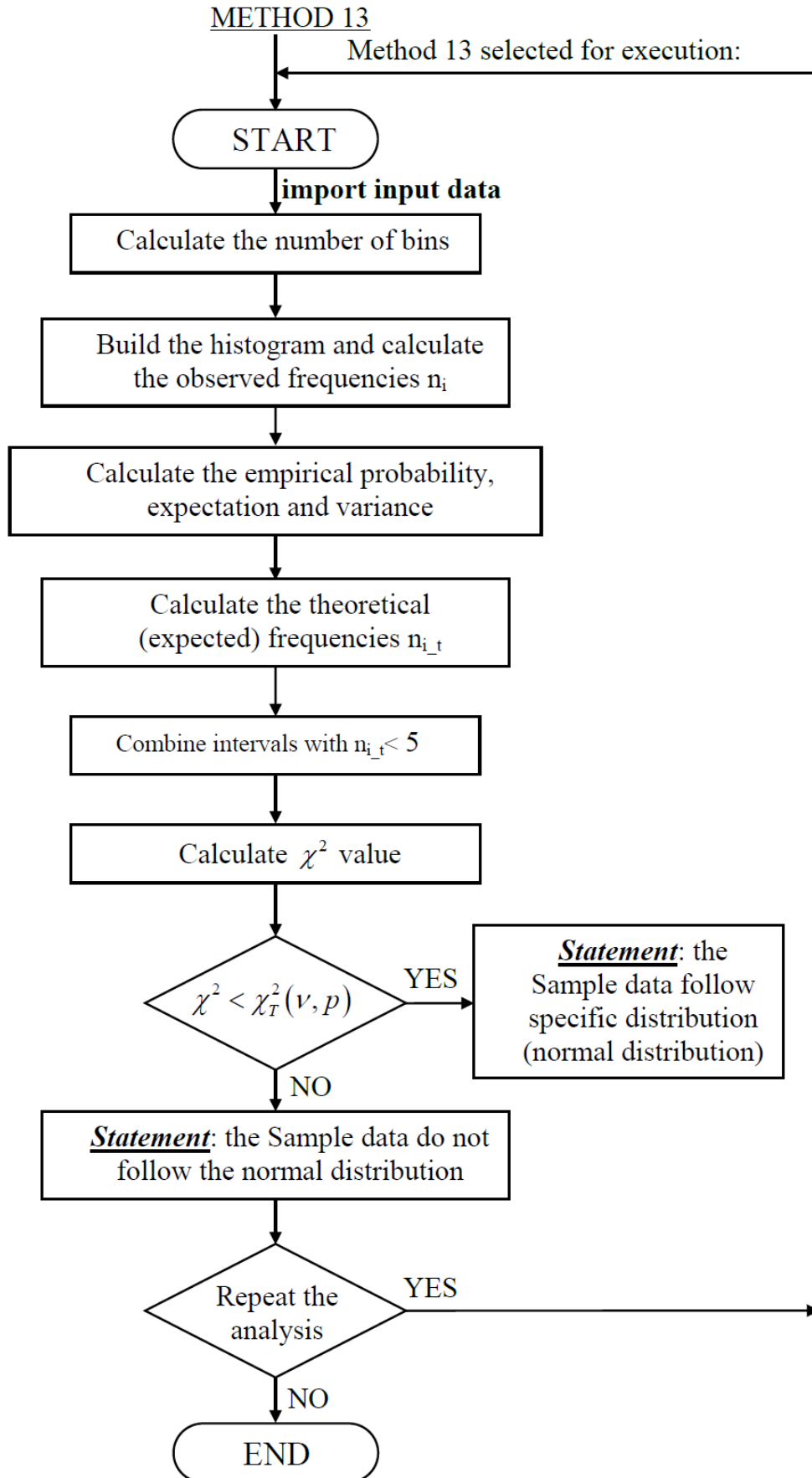
Repeat the analysis   YES →

NO

END

FIGURE 2. ALGORITHM BLOCK-DIAGRAM – METHOD 13

IV.   GENERAL PROGRAMMING STRUCTURES IMPLEMENTED IN SCRIPT DEVELOPMENT

The selection of a chosen method is considered an execution of one or several methods selected as per the user's needs by implementation of switch - case structure:

```
SWITCH switch_expr

     CASE case_expr,

          statement, ..., statement

       CASE {case_expr1, case_expr2, case_expr3,...}

          statement, ..., statement

     ...

       OTHERWISE,

         statement, ..., statement

     END
```

Related example is given below:

```
k = input('Choose a method k= ');

switch k

case 1

again = input ('Would you like to continue with another method?(yes/no): ','s');

if( ~ismember(again,{'y','Y','yes','Yes','YES'}) )

break

end

    …….

    Case 13

    end
```

Single method execution can be repeated as many times as needed for the purpose of analysis.

The input data (observations/measurements) is considered preliminary processed and saved in .txt files and further loaded upon execution as per the example given below:

```
L = importdata('C:\Users\USER\Desktop\PS\Lanchovy.txt');%form the input vector L for  the regression analysis%

W = importdata('C:\Users\USER\Desktop\PS\Wanchovy.txt');%form the outputvector W for the regression analysis%
```

If the data analyzed need to be processed to form the input-output data to feed further regression analysis, depending on the specifics of the method chosen it will be automatically processed upon execution with for loops and/or if statements (Example is given below):

```
n = r-1;

for i = 1:n

          GrR(i)=(L(i+1)-L(i))/deltat;%calculate the growth rate for time interval deltat%

          Lmean(i)=(L(i+1)+L(i))/2;%calculate the mean length reached for interval deltat%

End
```

and:

```
if (Cond_FF > maxval)

          disp('Cond FF>10 ==>')
```

```
        txt = sprintf('Cond FF>10 ==> \r\n');

        fprintf(fileID,txt);

        disp('Apply Regression analysis with normalized (standardized) variables')

        txt  =  sprintf('Apply  Regression  analysis  with  normalized  (standardized)
variables' \r\n');

        fprintf(fileID,txt);
```

The present development aims applicability and usage not only in MATLAB programming environment, but as a standalone executable application without requirements for MATLAB installation. Building a standalone application is possible with MATLAB Compiler, which can generate executable standalone files as command-line executable versions of the MATLAB code or complete applications that use MATLAB graphics and UIs designed with GUIDE. The designer of the applications can himself define the user inputs and choose how the results are presented using all the output formats which are supported by MATLAB (text, numeric, or graphical formats). End users of these applications do not need knowledge to operate in MATLAB programming environment unless they want to modify or design the applications themselves.

The program system developed is functioning as standalone application and records the results of the calculations in designated .txt file, which is organized within the program structure as given below:

```
filename =('C:\Users\USER\Desktop\PS\Results.txt');

    fileID = fopen(filename, 'a');%opens a designated .txt file to record the final and
interim results (selected as per their statistical and scientific significance for the
provision of the overall analysis goals%

..........................
disp('Regression coefficient b1=:')

        txt = sprintf(Regression coefficient b1=: %f \r\n' ,b1);%records the calculated
value of the regression coefficient in .txt file for further use and archive of analysis
results%

fprintf(fileID,txt)

    fclose(fileID)%closes the file upon completion of the method chosen for execution%
```

The user will be asked to enter table and critical values of variables on purpose – either to deliver statement of the type: "the model obtained is adequate/inadequate", or to be further used for calculation of parameters. Script example is given for F-test:

```
Fcr = input('Enter F-statistics table value Fcr=');

    txt = sprintf('Enter F-statistics table value Fcr=: %f \r\n' ,Fcr);

    fprintf(fileID,txt);


    if (Fn > Fcr)

        disp('The model is statistically significant and adequate')

        txt = sprintf('The model is statistically significant and adequate');

        fprintf(fileID,txt);

    else

        disp('the model is inadequate')

        txt = sprintf('the model is inadequate \r\n');

        fprintf(fileID,txt);

    end
```

The results recorded in the designated .txt file are chosen to be displayed in accordance with the parameters significance and meaning, interim results are only displayed and recorded if they are considered important indicators to prove statistical significance and accuracy of the parameters estimated. Example of data recorded in .txt file for archive or further use and processing of data, depending on the user needs is given below:

b. Statistical analysis

Correlation coefficient value R: 0.920405

F-test the overall model significance

Calculated F-statistics value Fl= 1.263613e+003

Enter critical (table) F-statistics value Fcr=: 3.450000e+000

The model is adequate

The program performance and accuracy is tested with experimental data, simulation data and solved numerical examples [4,13].

V.   CONCLUSIONS

The software developed delivers estimates of key growth and mortality parameters to support stock assessment procedure and analysis of historical and present data. It allows analysis of the overall model significance and delivers a statement for the model adequacy.

The program itself does not add computational error to the analysis results and works well with natural, standardized and log-transformed values of the variables under used for the provisions of analysis.

The methods can be repeatedly executed with new or amended data depending on the user's requirements and data availability.

In case that input data is available to feed few models, which deliver estimates of same stock/individual parameters, comparison of the estimates delivered might results in more realistic values to support the analysis.

The program is set as standalone application and it does not require MATLAB software installation to operate. In case the script needs further modifications as per the users demand it can be easily modified in MATLAB programming environment. The graphic representations dispose the same features as regular MATLAB graphs.

The results are recorded in a designated .txt file and can be further processed with other applications depending on the user needs.

REFERENCES

[1] Anderson, R. O., Neumann, R. M. 1996. Length, Weight, and Associated Structural Indices, in: B.E. Murphy and D.W. Willis, (Eds): Fisheries Techniques, 2nd edition, American Fisheries Society, Bethesda, Maryland: pp.447-481

[2] Belsley, D. A., Kuh, E., Welsch, R. E. 1980. The Condition Number. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley & Sons, New York.

[3] Blackwell, B. G., Brown, M. L., Willis, D. W. 2000. Relative Weight (Wr) Status and Current Use in Fisheries Assessment and Management, Reviews in Fisheries Science, 8(1): pp. 1–44

[4] Cadima, E.L. 2003. Fish Stock Assessment Manual. FAO Fisheries Technical Paper. No 393, Rome, FAO, 161p

[5] Cheney, W., Kincaid, D. 2008. Numerical Mathematics and Computing, 6th edition. Thomson Brooks/Cole

[6] Froese, R. 2006. Cube law, condition factor and weight-length relationships: history, meta-analysis and recommendations. Journal of Applied Ichthyology, 22: pp. 241-253.

[7] Genov, D., 2000. Modeling and Optimization of industrial processes Manual Lab, Varna: Technical university-Varna, 192p.

[8] Hahn, G. J. and Shapiro, S.S., 1967. Statistical Modeling in engineering, New York, London, Sydney: John Wiley and Sons Inc., 376p.

[9] NIST/SEMATECH e-Handbook of Statistical Methods, 2003, last updated: 10/30/2013:
    http://www.itl.nist.gov/div898/handbook/03.11.2017.

[10] Papoulis, A., 2002. Probability, Random Variables and Stochastic Processes (4th Edition). New York: McGraw-Hill Higher Education p. 168.

[11] Pesaran, M. H.. 2015. The Multicollinearity Problem. Time Series and Panel Data Econometrics. New York: Oxford University Press. pp. 67–72

[12] Snedecor G.W and Cochran W.G, 1989, Statistical Methods, 8th Edition, Iowa State University Press, 491p.

[13] Sparre P. and Venema S.C. 1998. Introduction to Tropical Fish Stock Assessment - Part 1: Manual. Rome: FAO Fish Tech. Pap., 306/1 (Rev.2), 407p

[14] Whetherall, J.A., Polovina, J. J, Ralston, S. 1987. Estimating Growth and Mortality in Steady-state Fish Stocks from Lenght-Frequency Data pp 53-74, In D. Pauly and G.R. Morgan (eds.) Length-based methods in Fisheries research. ICLARM Conference proceedings 13, 468p. Safat, Kuwait.