

Software Application for Electrical Distribution System Reliability Studies: Box-Jenkins Methodology.

Adejumobi Isaiah Adediji* and Olanipekun Ayorinde Joseph

Department of Electrical/Electronics Engineering, College of Engineering,
University of Agriculture, Abeokuta, Abeokuta, Nigeria.

*E-mail: engradejumobi@yahoo.com
Telephone: +2347033215455

ABSTRACT

The need for a fast and accurate technique to calculate the reliability of electrical distribution systems cannot be over emphasized. The duration of interruption in electricity distribution system is quite high and comes at random. In this paper, a software based on Box-Jenkins time series mathematical techniques, has been developed for fault predictions and reliability calculations on distribution systems. The program was compiled and molded using 6.0 version of Visual Basic Microsoft®. Three working modules were built into this computer program: the analysis module, the forecasting module, and reliability module. For each module, appropriate codes and subroutines were written. These modules were presented in visualizing forms for direct interaction by the user. Using the available six years (2001-2006) faults data from Dugbe Distribution System, Jericho Undertakings, Ibadan. The developed program was used to simulate faults for another ten years from which the overall reliability of the system was determined. The adequacy of the chosen mathematical model for fault prediction was determined by comparing the computed chi-square with Standard Statistical values. The developed software program is flexible and can be upgraded as data varies.

(Keywords: software development, distribution system, faults studies, time series, reliability calculation, visualization)

INTRODUCTION

Electricity is one of the most essential necessities of modern life, with all aspects of society becoming more dependent on it for successful functioning. Hence, the link between the source of electrical power, the consumer, and the

distribution system, assumes an ever more critical role.

The frequent power outages in Nigeria due to inefficient and unreliable distribution systems and poor management pose serious problems to consumers, and threats to the national economy. One of the major benefits of electrical distribution system fault forecasting is the use of such information to assess the reliability of the system. A system is said to be reliable if it meets up to a certain level of expectation, and is unreliable if otherwise. The reliability of a system can be defined as “the probability that the system uninterruptedly performs certain (accurately) specified functions during a stated interval of a life durable, on the condition that the system is used within a certain specified environment” (Klass and Jack,1998). In simpler terms, reliability can be defined as “the probability that the system will operate to an agreed level of performance for a specific period, subject to specified environmental conditions” (Cluley 1974). In other word, reliability is the probability of failure free operation for a specified time.

The importance of computer applications to power system analysis and control is becoming more imperative due to the increasing complexity of data processing and analysis. Some of the involved data are probabilistic and stochastic. The advantage of using computers for data processing and analysis lies in their high computational accuracy and great speed. It is not subject to fatigue and mistakes or human errors during voluminous and cumbersome computations. The use of computers therefore affords the time analyst the advantage of being able to work with more data at greater speeds and with far more accuracy than he can ever hope to achieve by manual computations (Stagg and El-Abiad,1968; Anderson, 2001). Therefore

for distribution system fault forecasting and reliability studies, the development of a reliable software program is desired.

A reliable electrical faults forecast is obtained only when an appropriate mathematical model has been established. The determination of an appropriate model is usually through statistical means. In this work, the three Box-Jenkins forecasting models for time series (George and Gwilyn, 1976; William, 1990; and Adejumobi, 2005) are considered. All time series techniques used in forecasting are based upon the fundamental assumption that past patterns will repeat in the future, thus the assumption on which Box-Jenkins methodology is based.

EXISTING SOFTWARE SOLUTIONS TO TIME SERIES ANALYSIS

There are a number of very useful statistical analysis packages that have time series analysis capabilities. These packages, though not developed solely for time series analysis, can if programmed correctly, compute the sample autocorrelation and sample partial autocorrelation functions and some can actually forecast. It is however almost impossible to find a single package that will do everything. Two of these packages are particularly useful for the analysis of the time series using the Box-Jenkins modeling methodology. They are (Adejumobi, 2005, Anderson, 2001):

- **Statistical Package for the Social Sciences (SPSS)**[®]. This is a widely used package, which contains programs for most commonly used statistical methods. It can compute the sample autocorrelation and the sample partial autocorrelation and can be programmed to carry out differencing of the data. It however does not forecast.
- **Statistical Analysis Package (SAS)**[®]. This is a very comprehensive package of statistical programs that is particularly useful for routine data management and manipulation. Users can also write their own programs or sub-routines using data handling techniques.

These two packages have one major setback. They were not developed solely for time series analysis and as such they require very deep knowledge from the users of the package to get meaningful results out of them.

A third option is the use of spreadsheets such as the Microsoft Excel[®], Lotus 123[®], Corel Perfect Expert[®], etc. Spreadsheets can be programmed to carry out basic manipulation of numerical data and as such are useful in data analysis such as the analysis of time series.

TIME SERIES ANALYSIS

A time series is a set of observation generated sequentially in time. (George & Gwily, 1976; William, 1990). The ordering of the series is usually through time, particularly in terms of some equally spaced intervals. A time series is said to be continuous when observations are made continuously in time, and it is said to be discrete if the set of series is discrete (Adejumobi, 2005). All time series techniques used in forecasting are based upon the fundamental assumption that past patterns will repeat in the future. For the purpose of this work, a Box-Jenkins methodology of Time series is considered for fault forecasting. Using this technique, the analyst does not arbitrarily pick a specific model for forecasting, but instead eliminates inappropriate models until he is left with the most suitable one. (George and Gwily, 1976 and Adejumobi, 2005). However, a good deal of experience and initiative judgment on the part of forecaster is often involved in the determination of the appropriate model. Three steps are involved in Box-Jenkins models development. The procedure includes model identification, coefficient estimation and diagnostic checking (Adejumobi, 2005)

Autocorrelation and Autocorrelation Functions: In constructing a model of the time series one needs to describe the relationship between a current observations and previous observations. This relationship is measured by the sample Autocorrelation defined as autocorrelation function, r_k (Adejumobi, 2005). The r_k at lag k can be written as (George & Gwily, 1976):

$$r_k = \frac{\sum_{t=1}^n [(z_t - \mu)(z_{t+k} - \mu)]}{\sum (z_t - \mu)^2} \quad (1)$$

where $Z_{t is}$ is the Observed value at interval t μ is the sample mean of observed value and,

$$\mu = \frac{1}{n} \sum_{t=1}^n z_t \quad (2)$$

n = number of observation, k = time lag and,

$$k = n/N \quad (3)$$

N is usually taken to be N = 4 for a fairly large data size (i.e., for n ≥ 36) (George & Gwily,1976). Two relationships are useful in describing time series behavior (Cluley,1974). The relationships are called the autoregressive (AR) and moving average (MA) models. The sample autocorrelation (r_k) is used to describe which of the two models is appropriate for a particular time series forecasting. For a process containing infinite a large number of parameter it may be necessary to combine both autoregressive and moving average terms. This combination is called mixed autoregressive moving average (ARMA) models (George and Gwilyn,1976).

Auto-Regressive Models: If a time series is best described by an autoregressive model, the autocorrelation exhibits decaying pattern of an exponential, oscillating sine wave (Adejumobi, 2005) . In this model, the current value of the series under study is expressed as a linear combination of previous values of the time series that explain the current observation and unexplained portion a_t. The autoregressive process of order P, (AR_(p)) is given as (George and Gwilyn,1976):

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} \cdot \cdot \cdot \cdot \phi_p Z_{t-p} + a_t \quad (4)$$

where $\Phi_1, \Phi_2, \dots, \Phi_p$ are parameter coefficients, or weighting coefficients and are usually determined, using the method of least square regression (Adejumobi, 2005)

Moving Average Models: When a time series is best described by a moving average model, the theoretical autocorrelation are noted by large spikes. The general qth order of moving average process, (MA(q)) is given as (George and Gwily,1976):

$$Z_t = a_t - \theta_1 a_{t-1} \cdot \cdot \cdot \cdot - \theta_q a_{t-q} \quad (5)$$

where again a_t, forecast error is given by:

$$Z_t - z_t = a_t = (\text{forecast} - \text{observed}) \quad (6)$$

For MA(1) and MA(2) models, the coefficient θ_1 and θ_2 are determined from autocorrelation function.

Mixed Auto-Regressive Moving Average Models: A natural extension of the autoregressive and moving average models is to combine both models to alleviate the above-mentioned problem. Such mixed process are referred to as Autoregressive-moving Average Models of orders p and q, or simply writer as ARMA (p,q). The process can be written as (Adejumobi, 2005):

$$Z_t = \phi_1 z_{t-1} + \cdot \cdot \cdot + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} \cdot \cdot \cdot - \theta_q a_{t-q} \quad (7)$$

It should be noted that the identification of a mixed ARMA model is not straightforward as in the case of either pure AR or MA model. However, whenever a mixed is present, both the sample autocorrelation and sample partial autocorrelation have a decaying pattern

Parameter Coefficients Estimation for Auto-Regressive And Moving Average Models: The estimation of coefficients for the above models has been described by George and Gwily (George and Gwily,1976). Most often the estimate requires iterative procedure (George and Gwilyn,1976 and Adejumobi, 2005).

Identifying The Orders Of AR, MA, and ARMA Models: In some cases, the sample autocorrelation described above may not indicate the appropriate order, most especially the autoregressive. Therefore, a second useful measure is the sample partial autocorrelation (SPAC) (George and Gwily,1976). This can be calculated for lags 1, 2, ..., k periods between time series values. The sample partial autocorrelation (A_{kk}) is defined as (Adejumobi, 2005):

$$A_{KK} = r_1 \quad (8)$$

for K = 1

$$A_{KK} = \frac{r_K - \sum_{j=1}^{r-1} A_{K-1,j} r_{K-j}}{1 - \sum_{j=1}^{r-1} A_{K-1,j} r_j}, \quad K = 2, 3, \dots, K \quad (9)$$

for $K = 1, 2, 3, \dots, \frac{n}{N}$

when n and N are as defined above.

The identification of a mixed ARMA of order (p, q) is not as straight forward as in the case of pure AR and MA models. When a mixed is present, both the sample autocorrelation and the sample partial autocorrelation have a decaying pattern (George and Gwily, 1976).

Test Of The Adequacy Of Chosen Model: The diagnostic checking is often carried out for model adequacy as follows:

Step 1: Use the tentative model to generate forecast values, Z_t

Step 2: Calculate the residuals, a_t ,

Step 3: Calculate the autocorrelation residuals $r_t(a_t)$

Step 4: Computer the chi-square statistic Q which is given as:

$$Q = n \sum_{k=1}^K r_k(a_t). \quad (10)$$

where n is the number of observations and Kis maximum lag.

Step 5: Calculate degree of freedom as $df = k-p-q$

Step 6: Test of adequacy:

The chi-square table is then checked for $\chi^2 [df]$. If $Q > \chi^2 [df]$ then the model is inadequate, otherwise is adequate.

Reliability Calculation: From the observed and forecast fault values, a mean failure rate, λ can be determined (Esan et al.,1993). The failure

rated, λ is defined as number of failure occurs per unit time (Cluley,1974; Adejumobi, 2005).

The reliability $R(t)$ is then calculated as (Cluley,1974; Endrenyi,1978; and Adejumobi, 2005):

$$R_{(t)} = \exp(-\lambda t) \quad (11)$$

Three phase fault will be considered in this work in that it represents total line failure. If $\lambda_{\text{year } k}$ (3Φ) is the failure rate for , three – phase faults in year k, then:

$$\lambda_{\text{year } k}(3\Phi) = \frac{\text{total } 3-\Phi \text{ faults in year } k}{\text{total nos of days in year}} \left(\frac{\text{faults}}{\text{days}} \right)$$

and hence, the reliability of the distribution system in year k is;

$$R_{(t)} = \exp(-\lambda_k \times t) \quad (11b)$$

NEED FOR SOFTWARE DEVELOPMENT FOR SYSTEM RELIABILITY CALCULATION

The advantage of developing fast and accurate calculation software for the above calculations lies in its high computational accuracy and great speed. The development of this program therefore affords the time series analyst the advantage of being able to work with more data at great speeds and with far more accuracy than manual computations (Stagg and El-Abiad, 1968; and Anderson, 2001). The Distribution System Reliability analysis program called Reliability Calculation Analysis Software (RCAS) was developed using the Microsoft Visual Basic® 6.0 programming language.

The Reliability Calculation Analysis Software (RCAS): The program was developed using the Microsoft Visual Basic® 6.0 programming language and its characteristic as mentioned earlier mentioned, the existing software packages for analyzing time series have one major setback. They were not developed solely for time series analysis and as such it requires a very good knowledge of the usage of the package to get meaningful result out of it. The analyst must understand programming techniques of the package to be able to use it. This setback is more pronounced in the case of analyzing electrical distribution system where the

time series analysis is just a means to an end. Therefore a computer program package that carries out time series analysis using the Box-Jenkins methodology was developed using Visual Basic® 6.0. This program contains three major modules: the Analysis module, the forecasting module, and the reliability module.

Microsoft Visual Basic® 6.0: The Microsoft Visual Basic® 6.0 is an object oriented programming language that is easy to learn. Unlike other programming languages such as C++ where one has to program every little detail, Visual Basic provides a high degree of simple automatic programming to your fingertips. Also, Visual Basic makes the creation of applications with Graphical User Interface easier and faster with the use of 'Controls'. Controls are the graphical tools that are used to build the user interface of a visual basic program. Controls are also known as objects (Reselman, Pruchniak, Peasley, and Smith, 1991).

Specification for Developed Computer Program: The program is required to analyze time series using the Box-Jenkins methodology and calculate system reliability. To achieve this, it is required to perform the following tasks:

- Accept time series (fault) data.
- Difference the data if requested by the user up to the second order.
- Plot a graph of the series i.e. a graph of number of faults against months.
- Compute the lag, SAC and the SPAC of the fault data.

- Plot a graph of the SAC and a graph of the SPAC against the Lag.
- Choose appropriate and diagnostics checking.
- Forecast.
- Print all generated graphs and tables.
- Compute the reliability of the system if requested of the user.

Program Architecture: The program is divided into three modules, which are the analysis module, forecasting module, and the reliability module and these modules relate to each other. Figure 1 shows diagrammatically the architecture of the developed program. The arrow indicates the possible directions of program flow. Which path is chosen depends on the user.

The Analysis Module: In analysis module, the Lag, SAC, and SPAC of the time series data are computed. The data can also be differenced in this module if satisfactory correct model has not been achieved. Figure 2a shows the flow chart followed in implementing the analysis modules. In this section, graphs of the original series, SAC and SPAC are plotted. The module consists of two forms. The first form is used to request data from a text file. The second form, which is the main form, consists of several intrinsic controls, which include command buttons, label controls, and text box controls. It also makes use of the grid control to display data and the MS Chart control to plot the graph. The Lag, Sac and SPAC were computed using the Equations (1), (8) and (9).

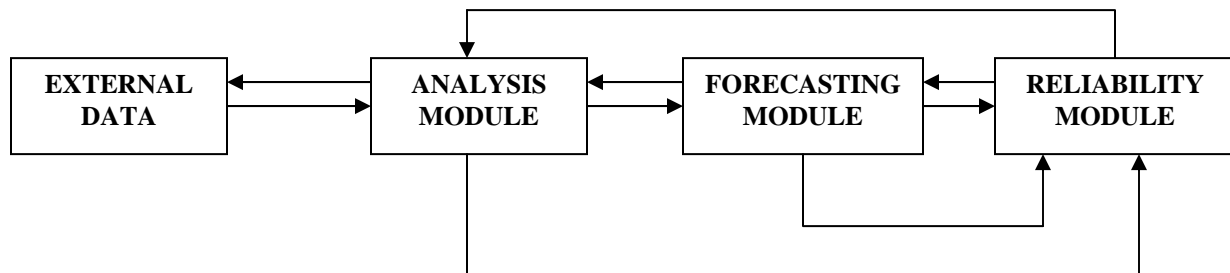


Figure 1: The Program Flow for the Developed Program.

Forecasting Module: This is the module where the actual forecasts are carried out using Equations (4) to (7). The diagnostic checking of the model using chi-square test is also done here. The module also consists of command button controls, label controls, and text box controls. A grid control is used to display the data while the MS Chart control is used to plot a graph of both observed values and the forecast values against months. Text box controls are used to allow the user to specify the model after which the forecast are calculated. The chi-square statistic is also computed using equation (10) and its value displayed. The program flow for the analysis module is shown in Figure 2b.

The Reliability Module: This is the final module and is also the simplest. It accepts input from the forecast module which is used to estimate the reliability of the system using the mathematical expressions (11) to (11b). All the modules can access each other and which module is displayed is entirely left to the user. Figure 2c shows the program flow for the forecasting module.

Codes listing: Without using the appropriate codes the program will introduce errors. The codes are constructed using Microsoft Visual Basic® 6.0 Professional Edition®. The codes are grouped into forms. The forms include data acquisition, analysis, forecasting form and reliability forms with their private sub commands embedded.

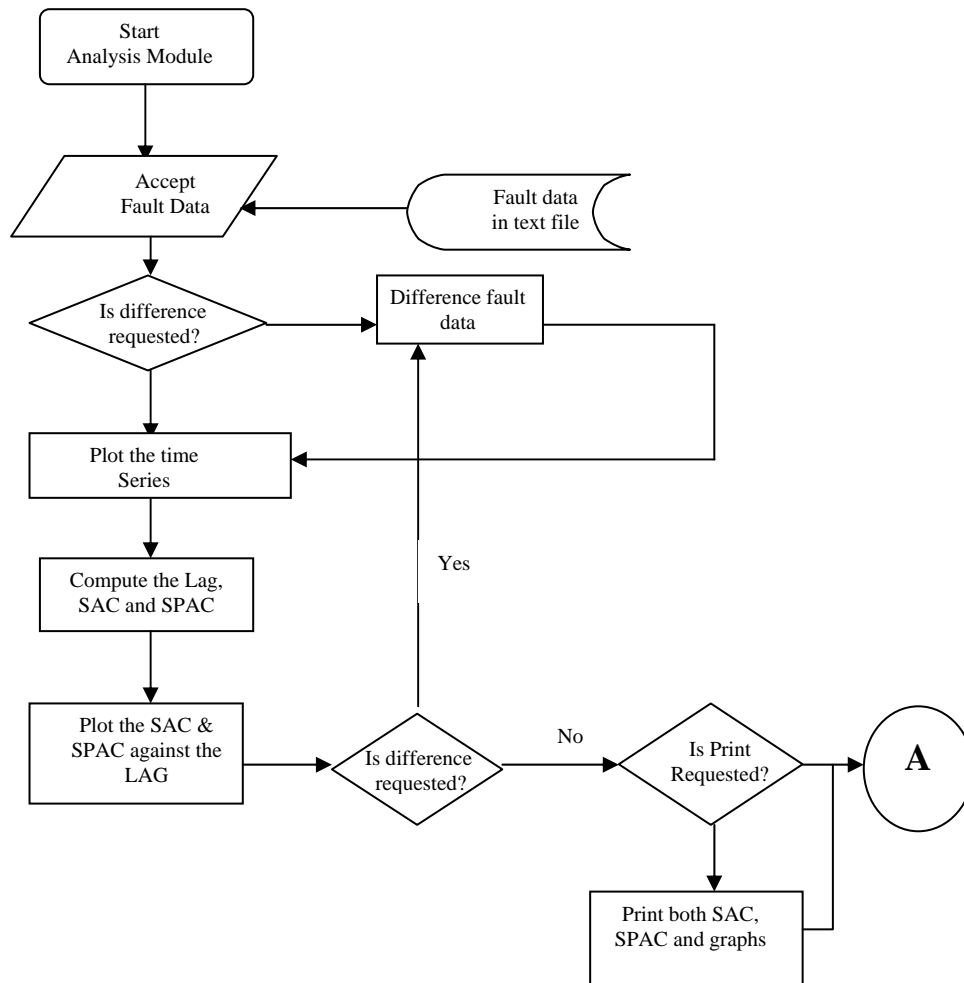


Figure 2 a: Program Flow for the Analysis Module to Determine Appropriate Model.

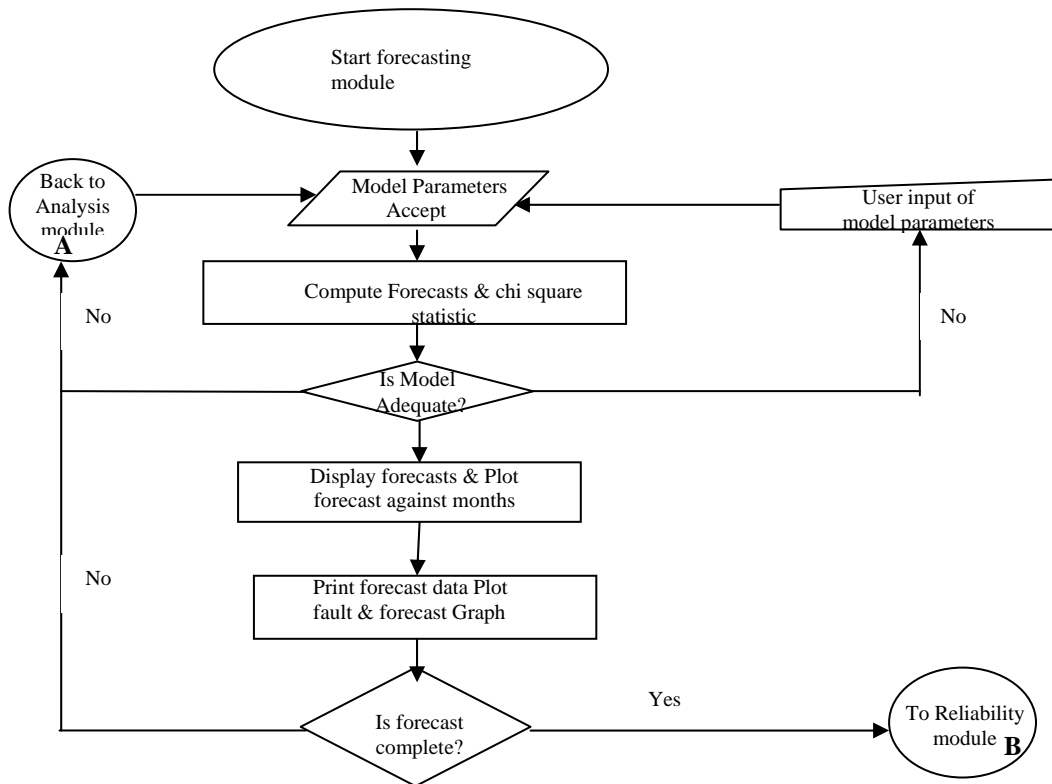


Figure 2b: Program Flow for the Forecasting Module.

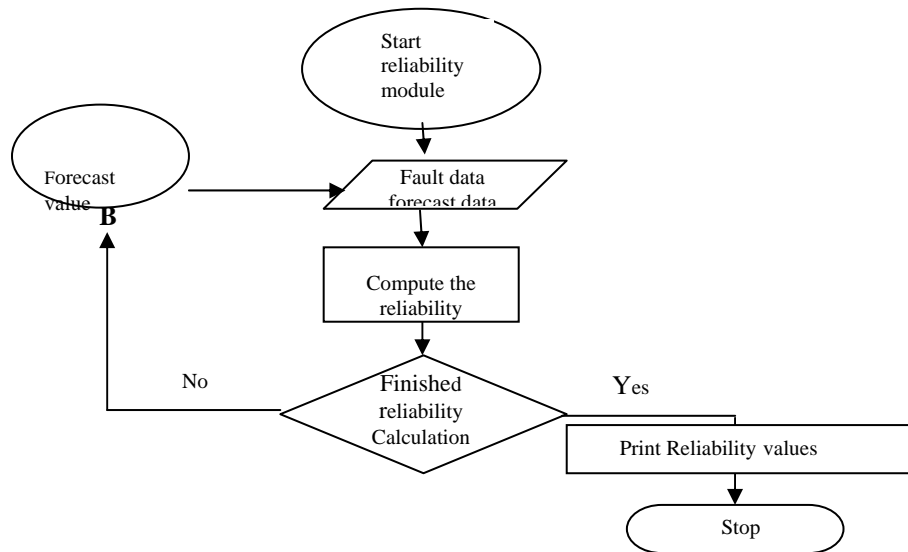


Figure 2c: Program Flow for the Reliability Calculation Module.

ANALYSIS AND RESULTS

In order to have direct interaction with the developed software and for easy visualization and interpretation of the results, the program package were developed into three majors forms: the analysis module, forecasting module and the reliability module. Each module contains coded subroutines of mathematical algorithms presented in Equations 1 to 11.

The developed program was tested using the three phase faults data collected from PHCN Dugbe Distribution systems, Jericho undertaking in Ibadan, Nigeria. Here, six years (2001-2006) of daily faults data from the logbooks as documented by the undertakings have been used as data base. Due to large data involved, these faults were summed up on monthly bases to cover each period six months as shown in Table1. This provided a database of 72 observations.

The sample autocorrelation and sample partial autocorrelation were computed to determine appropriate forecast models using the developed computer program. The adequacy of each chosen used in forecasting was determined at appropriate confidence level and degree of freedom, by comparing the calculated Chi-square with standard statistical Chi-square table. The observed and forecast values have been used to estimate systems reliabilities using expression (11b). The graphical views of the software analysis modules when the program was running are presented in Figures 3a, 3b, and 3c. Below also is a sample of the code listing during data preparations. There are also three other forms for codes listings for the three modules. The average overall reliability of the considered system was calculated to be about 40%.

Sample: CODE LISTING FOR THE DATA FORM (FORM 1)

```
Private Sub Command5_Click()
'Setting the properties of the common dialog box
  CmDbox1.DialogTitle = "Open Data File"
  CmDbox1.Filter = "Text Files (*.Txt) |*.Txt| "
  CmDbox1.FilterIndex = 0
  CmDbox1.ShowOpen
  FName = CmDbox1.FileName
  If FName = vbCncl Then Exit Sub
'Opening the data File and another text file for out put
of SAC/ SPAC data
  Open FName For Input As #1
  Open "c:\DSA\data.out" For Output As #2
  Input #1, N
'Setting the properties of the first flex grid object
  Form1.MSF1.Rows = N + 1
  Form1.MSF1.Cols = 3 + 1
  With Form1.MSF1
    .Row = 0
    .Col = 1
    .Text = "Data #"
    .Col = 2
    .Text = "Observation"
    .Col = 3
    .Text = "Model"
  End With
'Loading the data into the flex grid object on form1 and
unloading form2
  For i = 1 To N
    Input #1, w(i)
    Wsave(i) = w(i)
    With Form1.MSF1
      .Row = i
      .Col = 1
      .Text = i
      .Col = 2
      .Text = w(i)
    End With
  Next i
  Close (1)
  Msg = "Raw Data"
  Form1.Show
  Unload Me
End Sub
```

Table 1: Dugbe Distribution System, Jericho Undertaking, Ibadan: Three Phase Faults.

	2001	2002	2003	2004	2005	2006
JAN	100	21	106	55	43	45
FEB	90	23	89	59	76	71
MAR	82	21	132	85	56	65
APR	70	37	110	25	56	68
MAY	87	25	49	54	36	41
JUN	65	15	137	72	52	61
JUL	50	15	102	52	52	56
AUG	79	18	55	28	69	72
SEP	80	16	12	68	55	61
OCT	90	19	90	66	51	55
NOV	64	20	95	42	41	45
DEC	55	29	119	31	43	56

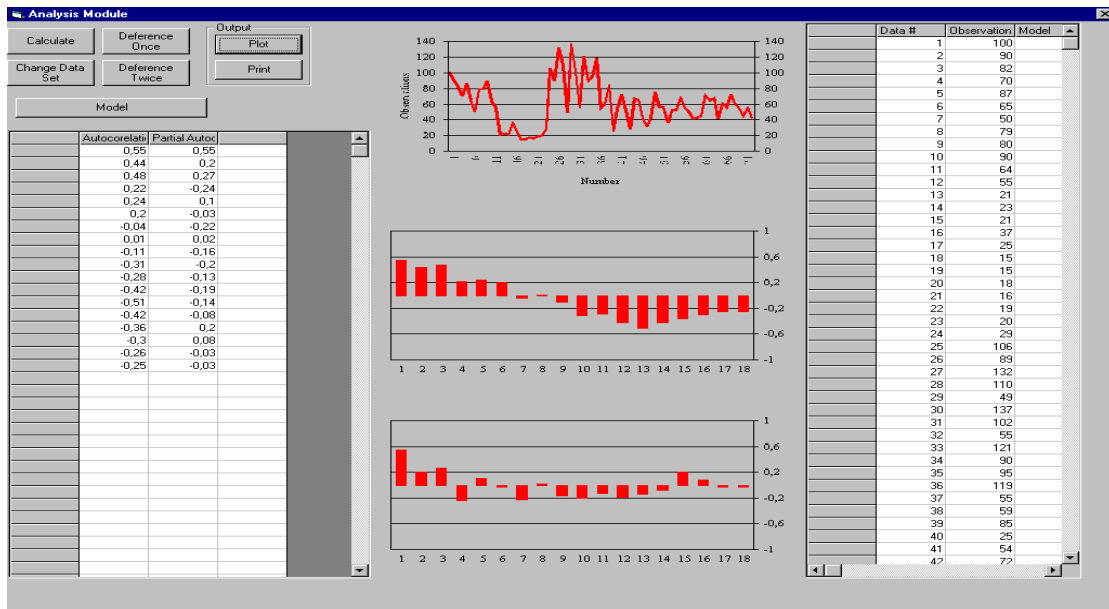


Figure 3a: A view of the Analysis Module while the Program is Running.

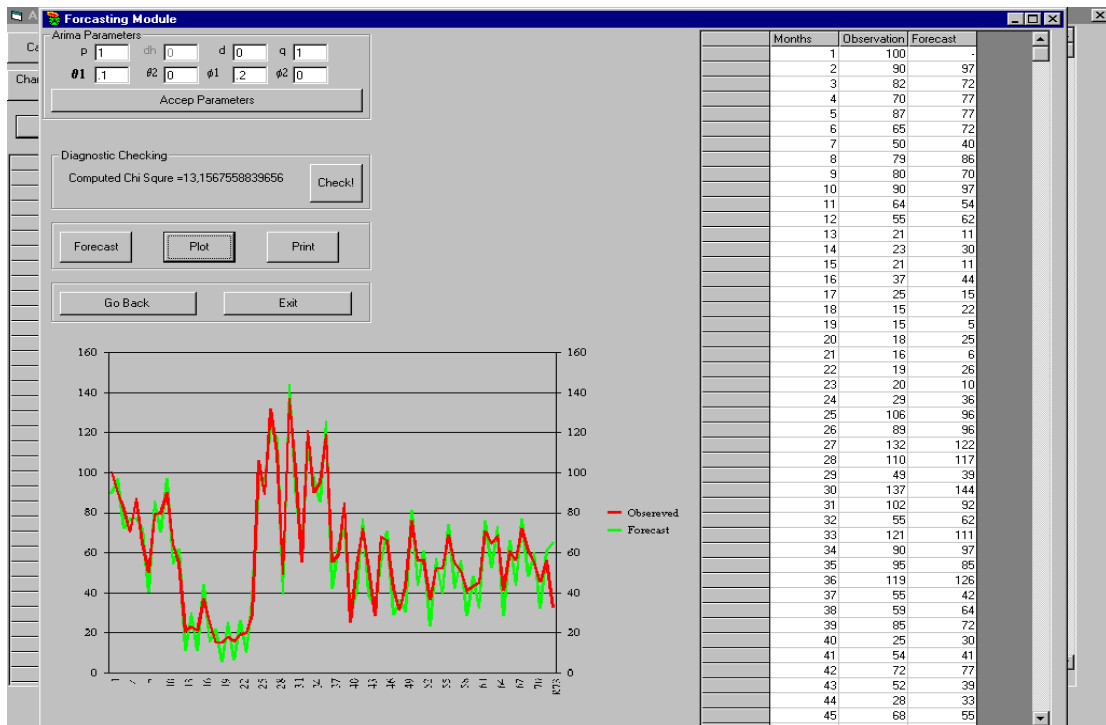


Figure 3b: A view of the Forecasting Module while the Program is Running.

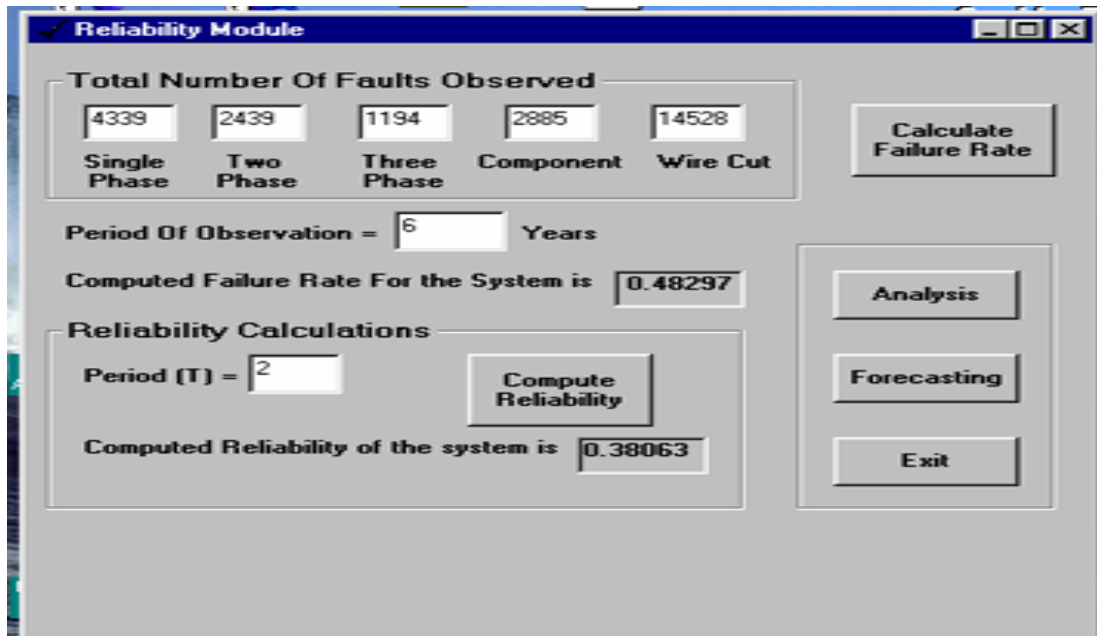


Figure 3c: A View of the Reliability Module while the Program is Running,

CONCLUSION

The application of computers to power system analysis has reduced considerably the complexity involved in the planning and control of power system management. The need to improve on the quality of electric power supply systems in Nigeria cannot be overemphasized. One of the major steps to achieve this is the systematic assessment of level of reliability in our distribution system.

To achieve a fast and accurate distribution system reliability calculations, a fast, accurate and flexible software called Reliability Calculation Analysis Software (RCAS) has been developed. The program made use of Box-Jenkins time series mathematical models and was compiled with Microsoft Visual Basic® 6.0 edition. The program is flexible to accommodate varying and large data and those data that can appear stochastically. The graphical visualization provision for the program makes it very interactive with the user.

The average overall reliability of the considered distribution system was about 40%. This program accepts any data where reliability of the system is to be determined.

REFERENCES

1. Adejumobi, I.A. 2005. "Assessment of Distribution System Reliability Using Time Series". *FUAJEET*. 4 (1):1-9.
2. Adejumobi, I.A. 2005. "Forecasting Models and Diagnostic Checking with Respect To Box-Jenkins Time Series Analysis". *Nig. J. Pure & Appl. Sci.* 20:1822-1832
3. Anderson, G.O. 2001." Computer Software Application to Determine Fault Level in Electrical Power System: A Case Study". *Int J Power Energy Syst.* 21(3):142-147.
4. Cluley, J.C. 1974. *Maintainability and Reliability*. 1st edition. Macmillan: London, UK.
5. Esan, A.A., Olufeagba, B.J., and Dada, J.O. 1993. "Reliability and Fault Analysis of Some Distribution Networks in Nigeria". *NSE Tech Trans.* 28(2):46-47.
6. Endrenyi, J. 1978. *Reliability Modeling in Electric Power Systems*. John Wiley and Sons: New York, NY.
7. George, B.F.P, and Gwily J.M. 1976. *Time Series Analysis for Forecasting and Control*. Holden-Day Inc.: Los Angeles, CA.

8. Klass, B.K and Jack, C.L.P. 1989. *System Reliability; Concept and Application*. Edward Arnold: London, UK.
9. Power Holding Company of Nigeria (PHCN). 2007 *Dugbe Distribution System, Jericho Undertaking, Ibadan, Nigeria: Faults Data Record Log Book (1997- 2006)*. PHCN: Lagos, Nigeria.
10. Reselman, B, Pruchniak, W, Peasley, A.R., and Smith, A.E. 1991. *Using Visual Basic 6.0, Macmillan Computer Publishing*. QUE Books: New York, NY.
11. Stagg, G.W. and El-Abiad, H.H .1968. *Computer Application in Power System Analysis*. McGraw Hill Inc.: New York, NY.
12. William, W.S.W. 1990. *Time Series Analysis*. Addison-Wesley Publication Co.: New York, NY.

SUGGESTED CITATION

Adejumobi, I.A. and A.J. Olanipekun. 2009. "Software Application for Electrical Distribution System Reliability Studies: Box-Jenkins Methodology". *Pacific Journal of Science and Technology*. 10(2):377-387.

 [Pacific Journal of Science and Technology](http://www.akamaiuniversity.us/PJST.htm)