TIME SERIES MODELING OF THE GLOBAL SOLAR

RADIATION DATA OF KNUST

BY

ARTHUR, ISAAC KOBINA EBO. BSc. (HONS.) MATHS

A Thesis submitted to the Department of Mathematics,

Kwame Nkrumah University of Science and Technology

In partial fulfillment of the requirement for the degree of

MASTER OF PHILOSOPHY

Department of Mathematics

[JUNE 2012]

CERTIFICATION

I hereby declare that this submission is my own work towards the MPhil degree and that , to the best of my knowledge it contains no material previously published by another person nor material which has been accepted for the award of any other degree of the university, except where due acknowledgement has been made in the text

Arthur Isaac K. E	-----------------------------	18-12-12
(PG5073810)	signature	Date

Certified by:

Mr. K. F. Darkwah	-----------------------------	-------------------------
(Supervisor)	signature	Date

Certified by:

Mr. K. F. Darkwah	-------------------------	-------------------------
(Head of Department)	signature	Date

ABSTRACT

In an attempt to help assess the solar energy resources potential of KNUST, a model was developed from data collected from the KNUST solar energy center to aid forecast future global radiation values. Many studies were carried out in order to develop this model. In this thesis, the Box Jenkins method was applied to the global solar radiation data from KNUST. By the application of time series analysis, a seasonal autoregressive moving average model that is ARIMA (0, 0, 1) (1, 0, 1) [12] was tentatively used to model the data. Ljung-Box statistic was used in diagnostic checking and it is shown that the model is adequate.

Keywords: Solar radiation, time series, Box-Jenkins method

TABLE OF CONTENT

CHAPTER TWO

LITERATURE REVIEW

CHAPTER THREE

METHODOLOGY

LIST OF TABLES.

LIST OF FIGURES

LIST OF ABBREVIATIONS

KNUST	Kwame Nkrumah University of Science and Technology
SEAL	solar energy application laboratory
H	global solar radiation
PV	Photovoltaic
MSD	meteorological services department
SE	solar energy
AR	autoregressive
MA	moving average
ARMA	autoregressive moving average
ARIMA	autoregressive integrated moving averages
ACF	Autocorrelation function
PACF	Partial autocorrelation function
SPSS	Statistical packages for social scientist
SAC	Sample autocorrelation
SD	Sunshine duration
WMO	World meteorological organization
BSRN	Baseline surface radiation network
ANNs	Artificial Neural Networks
RMSD	Root mean square deviation
r	correlation coefficient
MBE	Mean bias error
RMSE	Root mean square error
MPE	Mean percentage error
MABE	Mean absolute bias error

MAPE	mean absolute percentage error
GSR	Global solar radiation
MLP	Multi –layer perceptron
ISCCP	International satellite cloud climatology project
GEBA	Global energy balance archive
GPS	Global positioning system
DISSC	Direct Insolation simulation code
METSTAT	Meteorological and statistical
CSR	Climatological solar radiation
OLS	Ordinary least square
RSAC	Residual sample autocorrelation
RSPAC	Residual sample partial autocorrelation
AIC	Akaike information criterion
AICc	Akaike information criterion corrected
BIC	Bayesian Information Criterion
KPSS	Kwiatkowski, Phillips, schmidt and shin

DEDICATION

This work is dedicated to my children Kwesi Gyasi Arthur, Yaw Danquah Arthur and Kobby Arthur Jnr. It is also dedicated to my wife Mrs. Afia Abrafi Arthur for their love, care and prayer support.

ACKNOWLEDGEMENT

Now unto the king eternal, immortal, invisible the only wise God be honor and glory forever and ever. (1 Timothy 1:17). AMEN.

My next thanks goes to Mr. K. F. Darkwah the Head of mathematics Department who also served as my supervisor, for his patience, suggestions, contributions and time he had for me despite his busy schedules. Other thanks go to Mr. E. Harris, a lecturer of the mathematics department who served as my internal supervisor. Also to Degraft Owusu Ansah and Agyemang Isaac, my classmates, who helped me with the analysis of my data. Finally I would like to show my appreciation to Mr. I. A Edwin of the solar energy center who took time to explain to me all that goes on at the solar energy Centre.

CHAPTER ONE

INTRODUCTION

BACKGROUND TO THE STUDY

When solar power enters into the earth's atmosphere a part of the energy is lost by absorption by air molecules, clouds and particulate matter. Radiation that is not absorbed, and directly reaches the surface of the earth is called direct radiation. Absorbed radiation which reaches the ground is called diffuse radiation. Some of the radiation may reach a receiver after reflecting off the ground, and is called the albedo. The total radiation involved is called the global radiation. By definition it is the total quantity of short wave radiant energy emitted by the suns disc as well as that which is scattered diffusively by the atmosphere and cloud, passing through a unit area in the horizontal in a unit time.

Just as fossil fuel based energy industry relies on exploration and proven reserves for discovery and economic support of energy markets, so does renewable energy sector depends upon the assessment of resources for planning and selling their energy production technology. For solar-based renewable energy technologies such as solar thermal or photovoltaic conversion systems, the basic resource or fuel available is solar radiation. Assessment of the solar resource for these technologies is based upon measured data, where available. However, the sparse distribution in space, and particularly over time, of measured solar data leads to the use of modeled solar radiation as the basis for many engineering and economic decisions. Measured and modeled solar radiation has attendant uncertainties. Most solar radiation models rely on measured data for their development or validation, and often the uncertainty or accuracy of that measured data is unknown. Okundamiya and Nzeako (2011) upon research stated that Global solar radiation is an important parameter necessary for most ecological models and serves as input for different photovoltaic conversion system; hence is of economic importance to renewable energy alternative. Solar radiation reaching the earth's surface depends on the climatic condition of the specific site location, and this is essential for accurate prediction and design of a solar energy system. When global solar radiation is used to generate electrical energy for any specific site location, a provision should be made to forecast solar energy which will convert to electrical energy to recover the load demand, that is, the amount of solar energy for that place ought to be known. Technology for measuring global solar radiation is however costly and has instrumental hazards. Although solar radiation data are available in most meteorological stations, many stations in developing countries including Ghana suffer from a shortage concern of these data. Thus, alternative methods for estimating these data are therefore required. One of these methods is the use of empirical models. Accurate modeling depends on the quality and quantity of the measured data used and is a good tool for generating global solar radiation at locations where measured data are not available. According to Almorox and Hontoria (2003), Knowledge of local solar radiation is essential for many applications, including architectural design, solar energy systems and particularly for design methods, crop growth models and evapotranspiration estimates in the design of irrigation systems. Thus solar radiation is a required variable for the designers of solar energy systems. It is often provided as solar radiation maps, which is usually a preferable approach, more efficient and easy to handle. In order to build them, many procedures have been proposed. Most of them require knowing the solar radiation at many points spread wide across the region of interest. In addition, empirical, semi empirical, physical, neural networks, wavelets, fractals, etc. techniques must be assumed for the spatial prediction.

Unfortunately, solar radiation is still a scarcely sampled variable with respect to other environmental variables as temperature or precipitation, and there is usually a small set of stations with available radiation measurements. Again Solar radiation is said to be a primary driver for many physical, chemical and biological processes on the earth's surface, and complete and accurate solar radiation data at a specific region are of considerable significance for such research and application fields as architecture, industry, agriculture, environment, hydrology, agrology, meteorology, limnology, oceanography and ecology. Besides, solar radiation data are a fundamental input for solar energy applications such as photovoltaic systems for electricity generation, solar collectors for heating, solar air conditioning climate control in buildings and passive solar devices .it is of economic importance as a renewable energy alternative. Recently global solar radiation is being studied due to its importance in providing energy for Earth's climate system. The solar radiation reaching the Earth's surface depends upon climatic conditions of a location, which is essential to the prediction, and design of a solar energy system, Burari and Sambo (2001).Sunshine duration measurements have been made at many locations around the world for over a century, using different methods and instruments. It can be used to characterize the climate of a particular region hence applied in tourism. Furthermore, if solar radiation measurements are not available, sunshine duration data can be used as a proxy for the solar irradiance, which is a valuable quantity for agriculture, architects, and various solar energy applications (Velds 1992). A commonly used instrument for the measurement of sunshine duration is the Campbell–Stokes sunshine. Although considerable effort has been done to make use of solar energy efficiently from industrial revolution with the expectation that fossil fuels would run out in the future, only minimal resources have been directed towards forecasting incoming solar energy at ground level Vezirogluanddot, (2008). However, the necessity to have forecasting models which could optimize the integration of solar energy into electric grid is increasing as they gain recognition as an energetic source. Currently, the potential market of solar energy is huge. Its development is being supported by agreements in Kyoto protocol and by progressive series of regulations regarding green energy (feed-in tariff) established in several countries like Spain and Germany (lizianas et al., 2007). In the case of Spain, current legislation (Royal Decree 661/2007, 25th of May) allows promoters minimize investment risks and contribute to the development of solar energy. Rating and sizing of solar energy systems and performance analysis of such systems requires information on solar radiation at Earth's surface is available. Such information is also essential in many other applications such as crop growth models, evaporation–transpiration estimates and building comfort conditions. There is no doubt that measured data are the best source of obtaining information on solar radiation. But, the measurement of solar parameters is made only in meteorological stations. Therefore, it has been necessary to estimate solar radiation by theoretical models. The parameters affecting solar radiation can be categorized in astronomical, geographical, geometrical, physical and meteorological factors. Ertekin and Yaldiz, (1999). Knowledge of the local solar-radiation is essential for the proper design of building energy systems, solar energy systems and a good evaluation of thermal environment within buildings. The best database would be the long-term measured data at the site of the proposed solar system. However, the limited coverage of radiation measuring networks dictates the need for developing solar radiation models. Since the beam component (direct irradiance) is important in designing systems employing solar energy, such as high-temperature heat engines and high-intensity solar cells, emphasis is often put on modeling the beam component. An optimal use of the renewable energy needs its characterization and prediction in order to size detectors or to estimate the potential of power plants. Paoli1 and voyant (2010) .In terms of prediction, electricity suppliers are interested in various horizons to estimate the fossil fuel saving and to manage and dispatch the power plants installed. Artificial intelligence techniques are becoming more and more popular in the renewable energy domain and particularly for the prediction of meteorological data such as solar radiation, Thereby many research works have shown the ability of Artificial Neural Networks (ANNs) to predict times series of meteorological data.

In any solar energy conversion system, the knowledge of global solar radiation is extremely important for the optimal design and the prediction of the system performance. The best way of knowing the amount of global solar radiation at a site is to install pyranometers at many locations in the given region and look after their day-today maintenance and recording, which is a very costly exercise. The alternative approach is to correlate the global solar radiation with the meteorological parameters at the place where the data is collected. The resultant correlation may then be used for locations of similar meteorological and geographical characteristics at which solar data are not available. However, the necessity to have forecasting models which could optimize the integration of solar energy into electric grid is increasing as they gain recognition as an energetic source. Currently, the potential market of solar energy is huge. Information of local solar radiation is essential for many applications, including architectural design, solar energy systems and particularly for design. Unfortunately, for many developing countries, solar radiation measurements are not easily available due to the cost and maintenance and calibration requirements of the measuring equipment. Therefore, it is important to elaborate methods to estimate the solar radiation based on readily available meteorological data. A model is therefore developed by this thesis and used to predict or forecast future values. Data for the study was collected from the Solar Energy Application Laboratory (SEAL) at KNUST. The solar energy application laboratory (SEAL) was officially established in 1997 by the mechanical Engineering department. The type of data collected include; solar radiation, bright sunshine hour duration, relative humidity and air temperatures. KNUST measures solar radiation (global and diffuse) using Kipp and Zonen radiometers connected to a data logger. These instruments are generally classified as second-class equipment according to the World Meteorological Organization. The instruments are estimated to have a 5% margin of error. The MSD used to collect global solar radiation data until their instruments (Bellani Distillation Pyranometers, classified as third-class instrument with 15% error) broke down. As a result, they are currently only measuring the duration of bright sunshine using Campbell – Stokes sunshine recorders

1.1 PROBLEM STATEMENT

The MSD used to collect global solar radiation data until their instruments broke down. As a result, they are currently measuring the duration of bright sunshine only using Campbell-stokes sunshine recorders. Upon the establishment of the KNUST energy Centre, it is now the only institution in Ghana which measures solar radiation (global and diffuse).Currently the center has on record, radiation data from 1995 to 2004. Thus no data has been collected since 2005 and no alternative method has been put in place to generate future values

1.2 OBJECTIVES

The major aim of this thesis therefore is to use time series data analysis on the global solar radiation data of KNUST, specifically from January 1995 to December 2003. The following are the specific objectives of the study:

1. To model the average monthly global solar radiation data of KNUST using time series analysis

2. To predict the future global solar radiation values for KNUST.

1.3 METHODOLOGY

The KNUST energy center measures solar radiation (global and diffuse). Currently it has on record global radiation data from January 1995 to December 2004. There is no data record from 2005 to date. Moreover no alternative measure has been put in place to generate radiation values. By this thesis, time series ARIMA model is used to analyze the data.In the process the Box- Jenkins approach is used for the analysis of this data.Time series analysis of the datawas conducted with the help of R and SPSS.

The KNUST library and the internet were our main source of information.

1.4 JUSTIFICATION

Measured data are the best source of obtaining information on solar radiation. However the limited coverage of radiation measuring networks dictates the need for developing solar radiation models. Unfortunately, for many developing countries, including Ghana solar radiation measurements are not easily available due to the cost and maintenance and calibration requirements of the measuring equipment. It is therefore believed that this model when developed for the average monthly global solar radiation of KNUST will be useful in the following ways;

- As an alternative method of generating global solar radiation values for KNUST
- for accurate prediction and design of solar energy systems for KNUST
- to properly assess the solar energy resources potential of KNUST
- to serve as basis for engineering and economic decisions at KNUST

. It is therefore believed that a model built by this thesis will help the socio- economic development of Ghana.

1.5 STRUCTURE OF THE THESIS

Chapter one contains the background statement, problem statement, objectives and significance of the study. Chapter two reviews the relevant literature and theories of time series analysis withspecial explanation on the Box Jenkins method.

Chapter three deals with modeling and analysis of data collected. Here we apply the Box –Jenkinsmethod of modeling time series experiment.

Chapter four deals with the data analysis and results obtained from the Box Jenkins approach. Chapter five deals the conclusion and recommendations of the study.

CHAPTER TWO

LITERATURE REVIEW

2.0 INTRODUCTION

In the design and study of solar energy, information on solar radiation and its components at a given location is very essential. Solar radiation data are required by solar engineers, architects, agriculturistsand hydrologists for many applications such as solar heating, cooking, drying and interior illumination ofbuildings. For this purpose, in the past, several empirical correlations have been developed in order toestimate the solar radiation around the world. The main objective of this chapter is to review global solar radiation models. There are several formulae which relate global radiationto other climatic parameters such as sunshine hours, relative humidity and maximum temperature. Themost commonly used parameter for estimating global solar radiation is sunshine duration. Sunshine duration can be easily and reliably measured and data are widely available.

2.1 REVIEW

Raji et al (2012) stated that solar radiation is the driving force for a number of solar energy applications such as photovoltaic system for electricity generations. Hence, the determination of solar radiation data through various approaches becomes imperative. In this paper, the Box Jenkins method is applied to global solar radiation data for some western cities of Nigeria. Seasonal autoregressive model of order 1 ARIMA (1,1,0) (1,2,0) and auto regression of order 2 ARIMA (2,1,0) (2,2,0) were tentatively used to model the ground and satellite solar data for Lagos respectively. Likewise ARIMA (0,0,3) (1,1,0) and (2,2,0) (2,1,0) for Ibadan ground & satellite data. ARIMA (2,1,0)

(2,2,0) ARIMA (2,0,0) (2,1,0) for Akure ground & satellite data respectively. Ljung-Box statistic, Residual auto-correlation and partial residual autocorrelation plots were used in the diagnostic checking. Using the model it was shown that the models are quiet adequate

Akuffo et al (2003) also aimed at developing adequate,accurate and reliable solar and wind energy resources data and information and evaluation tools as an integral part of the nation's energy planning and policy framework. Specifically the study involved compilation and analysis of solar radiation data from KNUST and the MSD. For the solar radiation data, the MSD measurement was validated using the KNUST data. However MSD solar radiation was discontinued after 1988. Since then MSD has only measured sunshine duration. Thus the bright sunshine measurement was converted to monthly averages of daily solar irradiation using the Angstrom-page correlation. These estimates were then validated using KNUST data. The estimated average percentage error of the MSD estimates with respect to the KNUST data is 3% after validation. The MSD validated estimates were then compared with NASA satellite generated measurements. A set of time series covering a period of 10 years was then developed based on the data collected for all 22 synoptic stations across the country. Since the solar radiation estimates were made for a horizontal plane, their corresponding inclined plane equivalents were computed.

Almorox and Hontoria (2003) estimated the global solar radiation from sunshine hours for 16 meteorological stations in Spain, using only the relative duration of sunshine. They employed several equations including the original Angstrom–Prescott linear regression and the modified functions (quadratic, third degree, logarithmic and exponential functions). Estimated values were compared with measured values in terms of the coefficient of determination, standard error of the estimate and mean absolute error. All the models used fitted the data adequately such that each could be used to estimate global solar radiation from sunshine hours. It was also found that seasonal partitioning does not significantly improve the estimation of global radiation. It was finally concluded that the third degree models perform better than the other models, however the linear model was given much preference due to its greater simplicity and wider application

Okundamiya1 and Nzeako2 (2011) proposed a temperature-based model of monthly mean daily global solar radiation on horizontal surfaces for selected cities, representing the six geopolitical zones in Nigeria. Their model was based on linear regression theory and was computed using monthly mean daily data set for minimum and maximum ambient temperatures. The results of the three statistical indicators: Mean Bias Error (MBE), Root Mean Square Error (RMSE), and *t*-statistic (TS), performed on the model along with practical comparison of the estimated and observed data, validate the excellent performance accuracy of the proposed model.

Linares-Rodriguez, et al (2011) used four variables (total cloud cover, skin temperature, total column water vapor and total column ozone) from meteorological reanalysis to generate synthetic daily global solar radiation via artificial neural network (ANN) techniques. By using the reanalysis data as an alternative to the use of satellite imagery, the model was validated in Andalusia (Spain), using measured data for nine years from 83 ground stations spread over the region. The geographical location and the four meteorological variables were used as input data. Sixty five ground stations were used as training dataset and eighteen stations as independent dataset. The optimum network architecture yielded a root mean square error of 16.4% and a correlation coefficient of 94% for the testing stations. The results demonstrated the generalization capability of this approach over unseen data and its ability to produce accurate estimates and forecasts.

Zaharim et al (2009) stated that solar radiation is a primary driver for many physical, chemical and biological processes on the earth's surface, and also a driving force behind a number of solar energy applications such as photovoltaic systems for electricity generation, solar collectors for heating, solar air conditioning climate control in buildings and passive solar devices. Many studies were carried out in order to develop a method to estimate the solar radiation. In their paper, the Box-Jenkins method was applied to solar radiation data from Bangi, Malaysia. Nonseasonal autoregressive model of order 1, ARMA (1, 0), was tentatively used to model the data. Ljung-Box statistic was used in diagnostic checking and it was shown that the model was adequate.

Falayi et al (2008) developed a number of multilinear regression equations to predict the relationship between global solar radiations with one or more combinations of the weather parameters: for Iseyin Nigeria for five years (1995-1999). Using the Angstrom model as the base, other regression equations were developed by modifying Angstrom equation. The value of correlation coefficient (r) and value of Root Mean Square Error (RMSE), Mean Bias Error (MBE) and Mean Percentage Error (MPE) were determined for each equation. The equation with the highest value of r and least value of RMSE, MPE and MBE was used. It was concluded that the developed model could be used for estimating global solar radiation on horizontal surfaces.

Hinssen and knap (2006) evaluated two Pyranometric methods for the determination of sunshine duration (SD) from global irradiance measurements by means of summated sunshine seconds derived from Pyrheliometric measurements in combination with the WMO threshold of 120 W m^2 for the direct solar irradiance. The evaluation was performed using direct and global radiation measurements made at the Cabauw Baseline Surface Radiation Network (BSRN) site in the Netherlands for the period March 2005–February 2006. The "Slob algorithm" used 10-min mean and extreme values of the measured global irradiance and parameterized estimates of the direct and diffuse irradiance. The "correlation algorithm" directly relates SD to 10-min mean measurements of global irradiance. The cumulative Pyrheliometric SD for the mentioned period was 1429 h. Relative to this value, the Slob algorithm and correlation algorithm gave -72 h (-5%) and +8 h (+0.6%). It was concluded that, by the use of either algorithm, it was possible to estimate daily sums of SD from 10-min mean measurements of global irradiance with a typical uncertainty of 0.5–0.7 h day_1.

Isikwue et al (2012) proposed the coefficients for Angstrom - Prescott type of model for the estimation of global solar radiation in Makurdi, Nigeria using relative sunshine duration alongside the measured global solar radiation data (2001-2010). The model constants a and b obtained in their investigation for Makurdi were .138 and 0.488 respectively. The correlation coefficient of 89% (P=0.00) between the clear sky index and relative sunshine duration, as well as the coefficient of determination, R^2 of 79.5 obtained showed that the model fits the data very well. Hence, the very low mean standard error of 0.025 showed a good agreement between the measured and estimated global solar radiation. Consequently, the developed model in this work could be used with confidence for Makurdi, and other locations with similar climate conditions.

.Martı´n et al (2010) presented comparisons of statistical models based on time series applied to predict half daily values of global solar irradiance with a temporal horizon of 3 days. Half daily values consist of accumulated hourly global solar irradiance from solar rise to solar noon and from noon until dawn for each day. The models tested were autoregressive, neural networks and fuzzy logic models. Due to the fact that half daily solar irradiance time series is non-stationary, it has been necessary to transform it to two new stationary variables (clearness index and lost component) which are used as input of the predictive models. Improvement in terms of RMSD of the models essayed was compared against the model based on persistence. The validation process shows that all models essayed improve persistence. It was finally concluded that the best approach to forecast half daily values of solar irradiance was by neural network models with lost component as input,

Yorukoglu and Celik (2005) used Models such as the Angstrom–Prescott equation to estimate global solar radiation from sunshine duration. They investigated the goodness of the estimation of global solar radiation based on a set of statistical parameters such as R^2, RMSE, MBE, MABE, MPE and MAPE. A case study of the estimation models and global solar radiation estimation from sunshine duration was presented using five different models (linear, quadratic, cubic, logarithmic and exponential), which are the most common models used in the literature, based on 6 years long measured hourly global solar radiation data. The statistical parameters were clearly derived from the basics for both of the data sets, and the inconsistencies caused by this confusion and other factors were exposed.

Paulescu and Schlett (2003) tested six global solar irradiation models against data measured at three stations from Romania in the year 2000.They fitted an empirical global solar irradiation model which requires as input a single meteorological parameter associated with the mean daily cloud amount. The accuracy of the EIM was accepted and compared to that of parametric models, which need more than one meteorological datum as input. The main conclusion was that such simple empirical "local-models" are a useful alternative for the more complicated one. In addition, the details of obtaining EIM were presented as a driven-tool, which may serve as a guide to elaborate similar simple solar irradiation models in any other location as well.

Saffaripour and Mehrabian (2009): predicted the global solar radiation intensity in areas where meteorological stations do not exist and information on solar radiation cannot be obtained experimentally. The approach was to develop the multiple regression relations between the global solar radiation intensity (the dependent variable) and geographical, geometrical, astronomical and meteorological parameters (the independent variables). The independent variables used for this purpose were selected based on their ease of measurability outside the meteorological station and without expensive equipment. – Linear regression relations using one, two, three, four, five, six, and seven independent variables were developed to predict the global solar radiation intensity on horizontal surfaces. An advanced computer program based on least square analysis was used to obtain the regression coefficients. The relations having the highest correlation coefficients were selected. The study shows even when only one independent variable (declination angle) is used, the one variable regression relation predicts the global solar radiation with an accuracy that is satisfactory in most engineering applications.

ZohrabSamani1 (2009): introduced a Procedure to estimate the solar radiation and subsequently reference crop evapotranspiration using minimum climatological data. He described a modification to an original equation which uses maximum and minimum temperature to estimate solar radiation and reference crop evapotranspiration. The proposed modification allows for correcting the errors associated with indirect climatological parameters which affect the local temperature range. The proposed modification improves the accuracy of estimates of solar radiation from temperature.

Bulut and Buyukalaca (2006) developed a simple model for estimating the daily global radiation. The model was based on a trigonometric function, which has only one independent parameter, namely the day of the year. The model was tested for 68 locations in Turkey using the data measured during at least 10 years. It is seen that predictions from the model agree well with the long-term measured data. The predictions were also compared with the data available in literature for Turkey. It was expected that the model developed for daily global solar radiation will be useful to the designers of energy-related systems as well as to those who need to estimates the yearly variation of global solar-radiation for any specific location in Turkey.

Fooladmand H., (2012), estimated monthly sunshine hours based on minimum, maximum and mean monthly air temperature, and minimum, maximum and mean monthly air humidity, and the obtained equations were used for estimating monthly ETo based on the Penman-Monteith equation and the results were compared with the Penman-Monteith equation by using actual sunshine hours data. The results indicated that the derived equations for estimating sunshine hours had high accuracy for estimating monthly ETo with the Penman-Monteith equation. It was concluded that it is possible to estimate monthly ETo with Penman-Monteith equation without using the sunshine hours data.

Paolil et al (2011) presented an application of neural networks in the renewable energy domain. They developed a methodology for the daily prediction of global solar radiation on a horizontal surface using an ad-hoc time series preprocessing and a Multi-Layer Perceptron (MLP) in order to predict solar radiation at daily horizon. The results obtained were promising with nRMSE< 21% and RMSE < 998 Wh/m². The optimized MLP presented prediction similar to or even better than conventional methods such as ARIMA techniques, Bayesian inference, and Markov chains and k-

Nearest-Neighbors approximators. Moreover it was found that the data preprocessing approach can reduce significantly forecasting errors.

El-Sebaii1and Trabea 2 (2005) measured data of the monthly average global solar radiation on a horizontal surface H and the number of bright sunshine hour's n for five locations in Egypt was analyzed. The regression constants for the first, second and third order Angstroms type correlations for each location was calculated using the method of regression analysis. Comparisons between measured and calculated values of H are performed for the present locations. The values obtained for the RMSE, MBE and the MPE indicated that the second and third order Angstrom type correlations do not improve the accuracy of estimation of global radiation. Therefore, the measured data available for the selected locations are combined and a first order correlation has been proposed for all Egypt. Moreover, all Egypt correlation is extended to other Egyptian locations which are not included in the regression analysis.

Karoro et al (2011) used five years of global solar radiation data to estimate the monthly average of daily global solar irradiation on a horizontal surface based on a single parameter, sunshine hours, using the artificial neural network method. The five-year data was split into two parts in 2003–2006 and 2007-2008; the first part was used for training, and the latter was used for testing the neural network. Results obtained using the proposed model showed good agreement between the estimated and actual values of global solar irradiation. A correlation coefficient of 0.963 was obtained with a mean bias error of 0.055 MJ/m^2 and a root mean square error of 0.521 MJ/m^2. The single-parameter ANN model shows promise for estimating global solar irradiation at places where monitoring stations are not established and stations where we have one common parameter (sunshine hours).

Remund1 and Müller1 (2011) carried out an analysis based on long time series of global radiation with a duration of at least 40 years and the forecasts of global radiation till 2100.For the mean and most of the individual 25 examined sites the dimming for the period 1950 – 85 and the brightening [2, 3] for the period 1985 – 2009 is statistically significant. The negative trend during the dimming period is clearly stronger (approx. factor 2) than the positive trend during the brightening phase. The individual regions and groups of measurement sites show a great variety of different trends for the analyzed sub periods. The variation depending on the duration of measurement is also quite different from site to site. The future changes are relatively small. On an average the global radiation will decrease slightly. However, in the Mediterranean region the trend is positive (+ 2 – 3 % till 2100).

Assi and Jama (2010) employed a Number of mathematical correlations to predict the monthly average global solar radiation on horizontal using the sun hours as an input parameter. The study was carried out on two weather stations in the UAE, which are Abu Dhabi and Al Ain, using a daily weather data recorded for 13 years. The used correlations included the linear Angstrom-Prescott model and its derivations, namely, the second and third order correlations. Moreover, the single term exponential model, logarithmic model, linear logarithmic model and power model were all examined in this work. The performance of the aforementioned correlations as global solar radiation estimators was evaluated by comparing the predicted values with the measured values. Different statistical error tests were employed to examine the accuracy of the mathematical models. All fits performed well in Abu Dhabi and Al Ain, The linear Angstrom-Prescott model and the third order model performed the best for Abu Dhabi and Al Ain, respectively.

Mihalakakou et al (1999) describes a neural network approach for modeling and making short-term predictions on the total solar radiation time series. The future hourly values of total solar radiation for several years are predicted, by extracting knowledge from their past values, using feed forward back propagation neural networks. The results were tested using various sets of non-training measurements, the findings were very encouraging and the model was found able to simulate the future values of total solar radiation time series based on their past values. ``Multilag" output predictions were performed using the predicted values to the input database in order to model future total solar radiation values with sufficient accuracy. Furthermore, an autoregressive model is developed for analyzing and representing the total solar radiation time series. The predicted values of solar radiation were compared with the observed data series and it was found that the neural network approach leads to better predictions than the AR model.

Markvart et al (2005); describes a sizing procedure based on the observed time series of solar radiation. Using a simple geometrical construction, the sizing curve is determined as a superposition of contributions from individual climatic cycles of low daily solar radiation. Unlike the traditional methods based on loss-of-load probability, the reliability of supply enters in this method through the length of the time series of data used in the analysis. The method thus resembles techniques used in other branches of engineering where extreme values are considered as functions of certain recurrence intervals.

Wong and Chow (2001) considered solar radiation models for predicting the average daily and hourly global radiation, beam radiation and diffuse radiation. Seven models using the A ˚ angstrom– Prescott equation to predict the average daily global radiation with hours of sunshine are considered. The average daily global radiation for Hong

Kong (22.3N latitude, 114.3E longitude) is predicted. Estimations of monthly average hourly global radiation are discussed. Two parametric models are reviewed and used to predict the hourly irradiance of Hong Kong. Comparisons among model predictions with measured data are made.

.Bar-Severl and Kuangl (2004) describe the development and testing of a set of new and improved solar radiation pressure models for Global Positioning System (GPS) satellites that is based on four and one-half years of precise GPS orbital data. These empirical models show improved performance in both GPS orbit fit and prediction relative to the state-of-the-art models. Orbit-fit rms is improved 80 percent for Block IIR satellites and 24 percent for Block IIA satellites. Orbit-prediction accuracy improved58 percent for Block IIR satellites and 32 percent for Block IIA satellites. These new models are designated GSPM.04. It is shown that, after the implementation of these new models, Block IIA and Block IIR satellites perform about the same in orbit fit and in orbit prediction.1

Safari and Gasore (2009) estimate global solar radiation on horizontal surface using sunshine-based models. Angstrom-type polynomial of first and second order have been developed from long term records of monthly mean daily sunshine hour values and measured daily global solar radiation on horizontal surface at Kigali, Rwanda. Coefficients of those polynomials were derived using least square regression analysis. These coefficients were then used for the estimation of solar radiation in other places of Rwanda where measures of solar radiation do not exist but sunshine records are available

Myers D. R. (2003) Measurement and modeling of broadband and spectral terrestrial solar radiation is important for the evaluation and deployment of solar renewable energy systems. We discuss recent developments in the calibration of broadband solar radiometric instrumentation and improving broadband solar radiation measurement accuracy. An improved diffuse sky reference and radiometer calibration and characterization software and for outdoor pyranometers calibrations is outlined. Several broadband solar radiation model approaches, including some developed at the National Renewable Energy Laboratory, for estimating direct beam, total hemispherical and diffuse sky radiation are briefly reviewed. The latter include the Bird clear sky model for global, direct beam, and diffuse terrestrial solar radiation; the Direct Insolation Simulation Code (DISC) for estimating direct beam radiation from global measurements; and the METSTAT (Meteorological and Statistical) and Climatological Solar Radiation (CSR) models that estimate solar radiation from meteorological data. We conclude that currently the best model uncertainties are representative of the uncertainty in measured data.

Sun and Kok (2007) sought to assemble solar radiation model capable of producing long-term, steady-state radiation forecasts. While computational methods for estimating solar radiation intensity outside the earth's atmosphere are well established, ground level fluxes were difficult to predict, given that the incoming flux is considerably attenuated by passage through the atmosphere, both as a result of its composition and of cloud distribution. The model was based on historical daily overall global solar irradiation (DOI, kJ m-2d-1) data, recorded in the Canadian cities of Vancouver, Winnipeg, Montreal, and Halifax, as supplied by the Meteorological Service of Canada. Analyzed and decomposed through a Fourier transform procedure, a single city-specific set of multiple single-year DOIs yielded city-specific descriptors, from which new year-long DOI sequences could be synthesized. Statistical testing ensured that synthetic data sets were sufficiently similar to the historical data sets.

CHAPTER THREE

METHODOLOGY

3.0 INTRODUCTION

This chapter deals with the basic concepts on time series,stationary and non-stationary time series, ARIMA model, (autoregressive integrated moving average), principles of ARIMAModeling (Box and Jenkins 1976) and conclusion. The objective of the study wasachieved alongside with statistical softwares such as R and SPSS.

3.1 BASIC CONCEPTS OF TIME SERIES

Time series by definition is a collection of observations made sequentially according to the timeof their outcome. It is a sequence of data points, measured typically at successive times, spacedoften at uniform time intervals. Time series analysis comprises methods that attempt to understand the underlying context of the data points by making forecasts or predictions. It involves the use of a model to forecast future events based on known past events, hence one is able to forecast future data points before they are measured. A standard example in econometrics isthe opening price of a share of stock based on its past performance.

The term time series analysis is used to distinguish a problem, firstly from more ordinary data analysis problems (where there is no natural ordering of the context of individual observations), and secondly from spatial data analysis where there is a context data, observations relate to geographicallocations. There are additional possibilities in the form of space time models (often called spatial temporal analysis). A time series model will generally reflect the fact that observations close together in time will be more closely related than observations further apart.Inaddition, it often makes use of the natural one way ordering of time, so that values in a seriesfor a given time will be expressed as been derived from past values, rather than from future values. A time series plot may reveal various features of the data such as;

- Trend, which indicates a long term change in the mean level as well as the upward ordownward movement that characterizes a time series over a period of time. Thus trend reflects the long –run growth or decline in the time series.

- Periodicity shows pattern repeating in time variations. The periodic patterns in a timeseries complete themselves within a calendar year and are then repeated on a yearly basis.This is usually caused by such factors as weather and customs.

- Unusual features, refers to irregular fluctuations component of a time series, this is in line with the erratic movements that follow no recognizable or regular pattern. Many irregular fluctuations in a time series are caused by events that cannot be forecasted suchas earthquakes, wars, and hurricanes. these features could be modeled in an additive form as

$$x_t = m_t + s_t + \omega_t \qquad t = 0, 1, 2 \ldots n \qquad (3.1)$$

Where

m_t = is a trend and usually a slowly changing function of time,

s_t = is a periodically function of time and

ω_t = is the random noise component

a trend $m_t$can exist as linear, where

$$m_t = \beta_0 + \beta_1 \ t \quad (3.2)$$

quadratic where

$$m_t = \beta_0 + \beta_1 \ t + \beta_2 t^2 \quad (3.3)$$

or polynomial of degree $k \geq 1$ where

$$m_t = \beta_0 + \beta_1 \ t + \beta_2 t^2 \ + \ \ldots\ldots\ldots\ldots + \beta_k t^k \quad (3.4)$$

As shown by Box and Jenkins (1970) and further in Box and Jenkins (1976), models for time series data can have many forms and represent different stochastic processes such as the autoregressive(AR) models, the integrated (I) models, and the moving averages (MA) models. These three classes depend linearly on previous data points, hence a combination of these ideas produces Autoregressive moving average (ARMA) and Autoregressive integrated moving average (ARIMA) models, the use of autoregressive fractionally integrated moving average

(ARFIMA) model generalizes the former three. Among other types of non – linear time series models, also exist models that represent the changes of variance along time (heteroskedasticity). These models are called autoregressive conditional heteroskedasticity (ARCH) and this collection comprises a wide variety of representation such as;

(GARCH, EGARCH, FIGARCH, CGARCH, etc). Here changes in variability are related to, or predicted by, recent past values of the observed series.

3.2 STATIONARY AND NON STATIONARY TIME SERIES

There are two types of stationarity to be discussed under this heading namely; stationary andnon-stationary time series.

3.2.1 Stationary Time Series

A series is said to be stationary if the mean and autocovariance of the series do not depend on time. The theory behind ARMA estimation is based on stationary time series. Stationary series are made up of strict and weak stationary series.

A strictly stationary time series is one for which the probabilistic behavior of every collection of values

$\{x_{t1}, x_{t2}, \ldots\ldots\ldots \ldots\ldots\ldots\ldots\ldots \quad x_{tk}\}$

is identical to that of the time shifted set

$\{x_{t1+h}, x_{t2+h}, \ldots\ldots\ldots, \ldots, \ldots\ldots\ldots x_{tk+h}\}$.

That is,

$$P\{x_{t1} \le c_1, \ldots, x_{tk} \le c_k\} = P\{x_{t1+h} \le c_1, \ldots, x_{tk+h} \le c_k\} \qquad (3.5)$$

for all k = 1, 2, ..., all time points $t_1, t_2, \ldots, t_k$, all numbers $c_1, c_2, \ldots, c_k$, and all time shifts h = 0,±1,±2, ...

If a time series is strictly stationary, then all of the multivariate distribution functions for subsets of variables must agree with their counterparts in the shifted set for all values of the shift parameter *h*. For example, when *k* = 1, (3.5) implies that

$$P\{x_s \le c\} = P\{x_t \le c\} \qquad (3.6)$$

for any time points s and t.

A weakly stationary time series x_t, is a finite variance process such that

(i) The mean value function, μ_t, is constant and does not depend on time t, and

(ii) The covariance function, $\gamma(s, t)$, depends on s and t only through their difference $|s - t|$.

We shall use the term stationary to mean weak stationarity; however if a process is stationary in the strict sense, we will use the term strictly stationary.

Because the mean function, $E(x_t) = \mu_t$, of a stationary time series is independent of time t, we will write

$$\mu_t = \mu. \tag{3.7}$$

3.2.2 Non Stationary Time Series

Most of the time series we encounter are nonstationary. Any series that is not stationary is saidto be non-stationary. For a nonstationary series to be made stationary it has to go through the due process of the Box – Jenkins approach in order to be made stationary.

A simple non stationary time series model is given by

$$x_t = \mu_t + y_t \tag{3.8}$$

Where themean $\mu_t$is a function of time andy_t is a weakly stationary series with mean zero. A random noise process $x_t$defined as

$$x_t = x_{t-1} + \omega_t \tag{3.9}$$

Is a nonstationary series, where ω_t is a stationary random disturbance term

Assuming $x_0 = 0$ then for t= 1`

$x_1 = \omega_1$ and

$$x_2 = x_1 + \omega_2 = \omega_1 + \omega_2 \tag{3.10}$$

by successive substitution

$$x_t = \sum_{i=1}^{t} \omega_i \tag{3.11}$$

HenceE(x_t) = t μ and var (x_t) = t σ^2

since the mean and variance change with t, the process is said to be nonstationary .

3.3 **ARIMA MODEL** (AUTOREGRESSIVE INTEGRATED MOVING AVERAGE)

ARIMA models (autoregressive integrated moving average) are generalization of the simple ARmodel that uses three tools for modeling the serial correlation in the disturbance. It is called an integrated model because the stationary ARMA model that is fitted to the differenced data has to be summed or "integrated "to provide a model for the nonstationary data. A differencestationary series is said to be integrated and is denoted as I (*d*) where d is the order of integration. Suppose we difference'd' times, to make the series stationary then the process is integrated of order d or I (d) and it has d unit roots. This will be modeled as an ARIMA(p, d,q)

the order of integration is the number of unit roots contained in the series, or the number ofdifferencing operations it takes to make the series stationary. Each integration order correspondsto differencing the series being forecasted. A first –order integrated component means that theforecasted model is differenced once with (d= 1) of the original series. A second order integrated component means the forecasted model is differenced twice with (d = 2). The three tool for modeling the serial correlation as contained in ARIMA models are as follows;

- An autoregressive model of order p i.e. AR(p)
- Moving average term of order q i.e. MA(q)
- Autoregressive moving average i.e. ARMA (p,q)

3.3.1 AUTOREGRESSIVE MODEL AR (p)

Autoregressive model AR (p) is a type of random process which is often used to model and predict various types of natural phenomenon. The idea behind the autoregressive models is to explain the present value of the series. x_t, by a function of (p) past values such as $x_{t-1,}x_{t-2,}x_{t-3}$ .... x_{t-p} therefore an Autoregressive process of order p is written as

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2+} \dots + \varphi_p x_{t-p} + \omega_t \tag{3.12}$$

Where ω_t is the white noise such that ω_t ~ WN $(0,\sigma^2)$ and is uncorrelated. Since AR is autoregressive. Writing equation (3.12) in terms of the lag operator L,

$$x_t = (\varphi_1 L + \varphi_2 L^2 + \dots + \varphi_p L^p) x_t + \omega_t \tag{3.13}$$

Now using the backward shift operator on equation 3.12 we obtain

$$\omega_t = x_t - \varphi_1 x_{t-1} - \varphi_2 x_{t-2} \dots - \varphi_p x_{t-p} \tag{3.14}$$

this simplifies to

$$\omega_t = (1 - \varphi B - \varphi_2 B^2 .- \dots, - \varphi_p B^p) x_t \tag{3.15}$$

Suppose we let $\omega_t = \varphi(B) x_t$ then

$$x_t = \frac{1}{\varphi(B)}\omega_t$$

Thus if $\omega_t = \varphi(B)x_t$ then an AR of order P can be simplified as

$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \ldots\ldots - \ldots\ldots\ldots \varphi_p B^p \ldots\ldots\ldots\ldots$ (3.16)

by definition an AR or order (1) is defined as

$x_t = \varphi\, x_{t-1} + \omega_t$ (3.17)

Where$\omega_t \sim$ WN $(0,\sigma^2)$ and φ = constant

we define AR (2) as

$x_t = \varphi x_{t-1} + \varphi_2 x_{t-2} + \omega_t$ (3.18)

this continues with AR (3) up to AR of order (p) as in the case of equation 3.12

to test for stationarity of AR(1) let

$$x_t = \varphi x_{t-1} + \omega_t$$

Then $x_t = \varphi(\varphi x_{t-2} + \omega_{t-1}) + \omega_t$

ie$x_t = \varphi x_{t-1} + \omega_t$

$= \varphi^2 x_{t-2} + \varphi\omega_{t-1} + \omega_t$

This simplifies to

$\varphi^k x_{t-k} + \sum_{j=0}^{k-1} \varphi\omega_{t-j}$ (3.19)

The later simplifies to

$\varphi^k x_{t-k} = x_t - \sum_{j=0}^{k-1} \varphi\omega_{t-j}$ (3.20) as $k \to \infty$

then

$$\lim_{K\to\infty} E(x_t - \sum_{j=0}^{k-1} \varphi\omega_{t-j})^2$$

This reduces to

$$\lim_{k\to\infty} \varphi^{2k} E(x_{t-k}^2) = 0 \qquad (3.21)$$

If $|\varphi| < 1$ and the variance of x_t is bounded then

$$x_t = \sum_{j=0}^{\infty} \varphi^j \omega_{t-j} \qquad (3.22)$$

Where in the mean sense

$$\varphi_j = \begin{cases} \varphi^j & for\ \ j \geq 0 \\ 0 & for\ j < 0 \end{cases}$$

As a result AR(1) is stationary and also has linear process with mean

$E(x_t) = \sum_{j=0}^{\infty} \varphi^j E(\omega_{t-j}) = 0$ and the variance

$$\gamma(0) = \frac{\sigma^2}{1-\varphi^2}$$

3.3.2 MOVING AVERAGE MODEL -MA (q)

As an alternative to the autoregressive representation in which the x_t on the left-hand side of the equation are assumed to be combined linearly, the moving average model of order q, abbreviated as MA(q), assumes the white noise ω_t on the right-hand side of the defining equation are combined linearly to form the observed data.

Thus the moving average model of order q, *or* MA (q) model, is defined to be

$$x_t = \omega_t + \theta_1\omega_{t-1} + \ldots\ldots\ldots\ldots\ldots\ldots \theta_q\omega_{t-q} \tag{3.23}$$

Where there are q lags in the moving average and θ_1.. ………… θ_q ($\theta q \neq 0$) are parameters. The noise ω_t is assumed to be Gaussian white noise.

We may also write the MA (q) process in the equivalent form

$$x_t = \theta(B)\omega_t \tag{3.24}$$

Thus using the moving average operator we have

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots\ldots + \theta_q B^q \tag{3.25}$$

Unlike the autoregressive process, the moving average process is stationary for any values of the parameters θ_1 ……. ,……. θ_q

Considering the MA (1) model $x_t = \omega_t + \theta_1\omega_{t-1}$ then

$$\gamma(h) = \begin{cases} (1+\theta^2)\sigma_\omega^2, & h = 0 \\ \theta\sigma_\omega^2, & h = 1 \\ 0, & h > 1 \end{cases}$$

And the autocorrelation function is

$$\rho(h) = \begin{cases} \dfrac{\theta}{(1+\theta^2)}, & h = 1 \\ 0, & h > 1 \end{cases}$$

3.3.3 AUTOREGRESSIVE MOVING AVERAGEMODEL; ARMA (p, q)

ARMA model are combination of the AR and MA models and is called autoregressive moving average which is given by.

$$x_t = \varphi_1 x_{t-1} + \ldots + \varphi_p x_{t-p} + \omega_t + \theta_1 \omega_{t-1} + \cdots + \theta_q \omega_{t-q} \quad (3.26)$$

Where ω_t is a purely random process with mean zero and variance σ^2

Using the lag operator L, we can write an ARMA process as

$$\varphi(L) x_t = \theta(L) \omega_t \quad (3.27)$$

Where $\varphi(L)$ *and* $\theta(L)$ are polynomials of orders p and q respectively, defined as

$$\varphi(L) = (1 - \varphi_1 L - \varphi_2 L^2 - \ldots - \varphi_P L^P) \quad (3.28)$$

and

$$\theta(L) = (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^P) \quad (3.29)$$

For example the ARMA (1, 1) process is given by

$$x_t = \varphi_1 x_{t-1} + \omega_t + \theta_1 \omega_{t-1}$$

And in terms of the lag operator L this can be written as

$$(1 - \varphi L) x_t = (1 + \theta_1 L) \omega_t$$

Hence

$$x_t = \left(\frac{1+\theta_1 L}{1+\varphi L}\right) \omega_t \quad (3.30)$$

Since ω_t is a pure random process with variance σ^2 we obtain variance as

$$var(x_t, x_{t-1}) = \frac{(\varphi+\theta)(1+\varphi\theta)}{1-\varphi}\sigma^2 \tag{3.31}$$

Where

$$\rho(1) = \frac{cov(x_t, x_{t-1})}{var(x_t)} = \frac{(\varphi+\theta)(1+\varphi\theta)}{1+\theta^2+2\varphi\theta} \tag{3.32}$$

Successive values of $\rho(k)$ can be obtained from the recurrence relation

$$\rho(k) = \varphi\rho(i-1) \qquad for \quad k > 2$$

Thus the ACF for an ARMA (1, 1) process is such that the magnitude of ρ_1 depends on both φ $\quad and \quad$ θ

3.4 MULTIPLICATIVE SEASONAL ARIMA MODELS

We now introduce several modifications made to the ARIMA model to account for seasonal and nonstationary behavior. Often, the dependence on the past tends to occur most strongly at multiples of some underlying seasonal lag *s*. For example, with monthly economic data, there is a strong yearly component occurring at lags that are multiples of s = 12, because of the strong connections of all activity to the calendar year. Data taken quarterly will exhibit the yearly repetitive period at s = 4 quarters. Natural phenomena such as solar radiation also have strong components corresponding to seasons. Hence, the natural variability of many physical, biological, and economic processes tends to match with seasonal fluctuations. Because of this,it is appropriate to introduce autoregressive and moving average polynomials that identify with the seasonal lags. The resulting pure seasonal autoregressive moving average model, say,

ARMA (*P*, *Q*) *s*, then takes the form

$$\Phi_p\,(B^S)x_t = \Theta_Q(B^S)\omega_t \tag{3.33}$$

with the following definition. The operators

$$\Phi_p\,(B^S) = 1 - \Phi_{\,1}B^S - \Phi_{\,2}B^{2S} - \cdots - \Phi_P B^{PS} \tag{3.34}$$

and

$$\Theta_Q(B^S) = 1 + \Theta_{\,1}B^{\,S} + \Theta_{\,2}B^{2S} + \cdots + \Theta_{\,Q}B^{QS} \tag{3.35}$$

arethe seasonal autoregressive operator and the seasonal moving average operator of orders P and Q, respectively, with seasonal period s.

Analogous to the properties of nonseasonal ARMA models, the pure seasonal ARMA $(P,Q)s$ is causal only when the roots of $\Phi_P(z^{\,S})$ lie outside the unit circle, and it is invertible only when the roots of $\Theta_Q(z^S)$ lie outside the unit circle.

In general, we can combine the seasonal and nonseasonal operators into a multiplicative seasonal autoregressive moving average model, denoted by

ARMA$(p,\ q) \times (P,Q)s$, and write

$$\Phi_p(B^s)\varphi(B)x_t = \Theta_Q(B^s)\theta(B)\omega_t \tag{3.36}$$

or

$$x_t = x_{t-1} + x_{t-12} + \cdots \quad + \omega_t + \theta\omega_{t-1} + \Theta\omega_{t-12} + \cdots \tag{3.37}$$

as the overall model.

Although the diagnostic properties in Table 3.1 are not strictly true for the overall mixed model, the behavior of the ACF and PACF tends to show rough patterns of the indicated form. In fact, for mixed models, we tend to see a mixture of the facts listed in Tables 3.1 and 3.2. In fitting such models, focusing on the seasonal autoregressive and moving average components first generally leads to more satisfactory results.

Selecting the appropriate model for a given set of data from all of those represented by the general form is a daunting task, and we usually think first in terms of finding difference operators that produce a roughly stationary series and then in terms of finding a set of simple autoregressive moving average or multiplicative seasonal ARMA to fit the resulting residual series. Differencing operations are applied first, and then the residuals are constructed from a series of reduced length. Next, the ACF and the PACF of these residuals are evaluated. Peaks that appear in these functions can often be eliminated by fitting an autoregressive or moving average component in accordance with the general properties of Tables 3.1 and 3.2. In considering whether the model is satisfactory, diagnostic techniques discussed is applied.

3.5 PRINCIPLES OF ARIMA MODELING (BOX JENKINS 1976)

Box Jenkins forecasting models are based on statistical concepts and principles and are able tomodel a wide spectrum of time series behavior. It has a large class of models to choose fromand a systematic approach for identifying the correct model form. There are both statistical tests for verifying model validity and statistical measures of forecast uncertainty. In contrast,traditional forecasting models offer a limited number of models relative to the complex behavior of many time series with little in a way of guidelines and statistical tests for verifying the validityof the selected model. It consists of a four step iterative procedure as follows;

1. Model identification

2. Model fitting

3. Model diagnostics and

4. Forecasting

The four iterative steps are not straight forward but are embodied in a continuous flow chart depending on the set of data one is dealing with or handling.

See figure 3.1 below for the chart of the box-Jenkins modeling approach for an ARIMA model

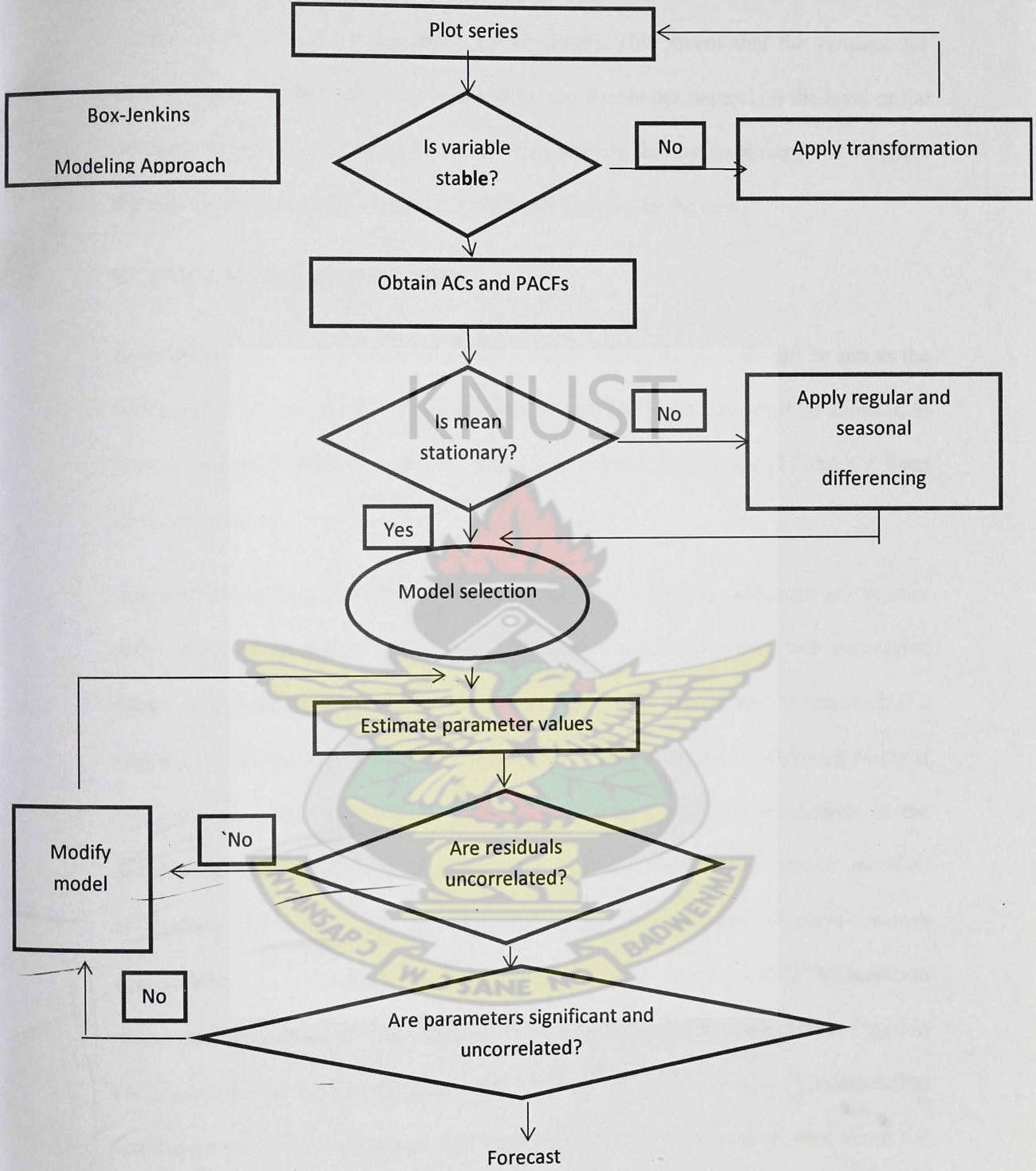


Figure 3.1 showing Box –Jenkins approach for ARIMA modeling

Referring to the above chart, it should be noted that the variance of the errors of the underlying model must be invariant (i.e. constant). This means that the variance for each subgroup of data used must be the same and should not depend on the level or the point in time. If this is violated then one can remedy this by stabilizing the variance and also making sure,there are no deterministic patterns in the data.

3.5.1 MODEL IDENTIFICATION

A Preliminary Box-Jenkins analysis with a plot of the initial data should be run as the starting point in determining an appropriate model. The input data must be adjusted to form a stationary series, one whose values vary more or less uniformly about a fixed level over time.

Apparent trends can be adjusted by having the model apply a technique of 'regular differencing' a process of computing the difference between every two successive values, computing a differenced series which has overall trend behavior removed. If a single differencing does not achieve stationarity, it may be repeated, although rarely if ever, are more than two regular differencing required. Where irregularities in the differenced series continue to be displayed, log or inverse functions can be specified to stabilize the series such that the remaining residual plot displays values approaching zero and without any pattern. Given a set of time series data like radiation under consideration, one can calculate the mean, variance, autocorrelation function (ACF), and partial autocorrelation function (PACF) of the time series. This calculation enables one to look at estimated ACF and PACF which gives one an idea about the correlation between the observations, indicating the sub-group of models to be entertained. This process is done by looking at the cut-offs in the ACF and PACF.

At the identification stage for this radiation data, we try to match the estimated ACF and PACF with the theoretical ACF and PACF as a guide for tentative model selection, but the final decision cannot be made until the model is estimated and diagnosed.

Table 3.1 Behavior of the ACF and PACF for Causal andInvertible ARMA Models

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts off after Lag q	Tails off
PACF	Cuts off after Lag P	Tails off	Tails off

Judge (1985) points out that when the PACF has a cut-off at p while the ACF tails off, it gives us an autoregressive of order p (AR (p)). If the ACF has a cut off at q while the PACF tapers off, it gives a moving average of order q (MA (q)). However, when both ACF and PACF tail off, it suggests the use of the autoregressive moving –average of order p and q ARMA (p,q) .

Table 3.2 Behavior of the ACF and PACF for Causal and Invertible Pure Seasonal ARMA Models

	AR(P)s	MA(Q)s	ARMA(P,Q)s
ACF	Tails off at lags ks, K=1,2,,………	Cuts off after Lag Qs	Tails off at lags ks
PACF	Cuts off after Lag Ps	Tails off at lags ks K=1,2, ………..	Tails off at Lags ks

Sometimes the ACF doesn't die out quickly, which may suggest that our stochastic process is nonstationary. This situation suggests the use of the ARMA (p, d, q) to difference the data (d) times, once or twice, until stationarity is obtained. This is the error term, equivalent to pure, white noise.

A good autoregressive model of order p (AR (p) has to be stationary, and a good moving average model of order q (MA (q)) has to be invertible. Invertibility and stationarity will give a constant mean, variance, and covariance. Anderson (1976) Chatfield (1984) and judge (1985) pointed out that by using what is called Wold's decomposition. In their submission , they pointed out that it is possible to show how the AR and MA processes are equivalent, thus causing one to expect that whenever a low – order –model of one type adequately explains a series, so should a higher-order model of the other. This expectation is valid only if the sum of the coefficients is less than one. Nevertheless, the principle of parsimony requires the model builder to choose the low –order-model, where the smallest possible number of parameters is employed for adequate representation.

Finally, quality of the coefficients has to meet two requirements. They must be statistically significant, and the correlation between the coefficients must be less than 0.9. The estimated ARIMA model has to have a significant t statistic for each coefficient of the estimated model

The correlation matrix measures the correlation between the estimated coefficients. The coefficients of the ARIMA model are correlated. however, if the absolute correlation coefficients between the two estimated ARIMA coefficients is 0.9 or more , such a coefficient value may suggest that the estimated coefficient are unstable and of poor quality. Under this condition the estimate could be inappropriate for future time periods, unless the behavior of future observations is the same as the behavior of a given realization.

3.5.2 MODEL FITING

Model fitting consists of finding the best possible estimates for the parameters of the tentatively identified model. In this stage, methods of estimation such as the method of moments, least-squares estimators and maximum likelihood estimators are considered to estimate the parameters.

A Box-Jenkins model is considered not invertible, if the weights placed on the past \z-observations when expressing z, as a function of these observations do not decline as we move further into the past. A model which is invertible on the other hand, implies that these weights do decline. Intuitively , this condition should hold , since it seems only logical that a recent observation should count more heavily than a more distantly past observation. The condition of stationary and invertibility implies that the parameters used in the model under consideration satisfy certain criteria. When we obtain the final least squares point estimates of the parameters in our model, we should verify that these point estimates satisfy the stationarity and invertibility conditions. The model will be considered inadequate if those conditions are not met.

3.5.3 MODEL DIAGNOSTICS

In model diagnostics, various diagnostics such as the method of autocorrelation of the residuals and the Ljung-Box-Pierce statistics are used to check the adequacy of the tentatively identified model. If the model is found to be inappropriate, we should return back to model identification and cycle through the steps until, ideally, an acceptable model is found. In order to achieve an acceptable model we test whether the estimated model conforms to the specifications of a stationary univariate process. In particular, the residuals should be independent from each other and constant in mean and variance over time. (Plotting the mean and variance of residuals over time and performing a Ljung-Box test or plotting autocorrelation and partial autocorrelation of the residuals are helpful to identify misspecification.) If the estimation is inadequate, we have to return to step one and attempt to build a better model which can be used to forecast future time series values.

The following are used in selecting the best or candid model;

a. Correlogram of the residuals

b. Normality test of the residuals; where under the normality test we consider the following

i. Histogram

ii. Swilk (Shapiro-wilk)

iii. Sktest (skewness-kurtosis)

c. Akaike information criterion (AIC) given by the relation

$$AIC= 2K \ln (L) \tag{3.38}$$

Where K is the number of parameters in the model, and L is the maximized value of the likelihood function for the estimated model.

d. Normalized Bayesian information criterion being one of the model selection criterions indicates that the model with the least NBIC value is selected among the other proposed models.

The ACF of the residuals can be examined in two ways. First, the ACF can be scanned to see if any individual coefficients fall outside some specified confidence interval around zero.

Approximate confidence intervals can be computed. The Correlogram of the true residuals (which are unknown) is such that r_k is normally distributed with mean

$$E(r_k) = 0 \tag{3.39}$$

And variance

$$\mathrm{Var}(r_k) = \frac{1}{N} \tag{3.40}$$

where (r_k) is the autocorrelation coefficient of the ARMA residuals at lag k. The appropriate confidence interval is found by referring to a normal distribution (CDF). For example, the 0.975 probability point of the standard normal distribution is 1.963. The 95% confidence interval for r_k is therefore ± 1.96 for the 99% confidence interval, the 0.995 probability point of the normal CDF is 2.57. The 99% CI is therefore± 2.57. anr_k outside this CI is evidence that the model residuals are not random. A subtle point that should be mentioned is that the Correlogram of the estimated residuals of a fitted ARMA model has somewhat different properties than the Correlogram of the true residuals which are unknown because the true model is unknown . A different approach to evaluating the randomness of the ARMA residuals is to look at the ACF 'as a whole' rather than at the individual ' r_k separately (Chatfield, 2004) . The test is called the portmanteau Lack of fit test and the test statistic is

$$Q = N \sum_{k-1}^{r} r_k^2 \tag{3.41}$$

This statistics is referred to as the *portmanteau statistic,* or 'Q' *statistic.* The Q statistic, computed from the lowest Kautocorrelations, says at lags k= 1, 2, 3 . . . . . . 20, follows a chi-square distribution with (k-p-q) degrees of freedom , where p and q are the AR and MA orders of the model and N is the length of the time series. If the computed Q exceeds the value from the chi-square table for some specified significance level, the null hypothesis that the series of autocorrelation represents a random series is rejected at that level. The p-*value* gives the probability of exceeding the computed Q by chance alone, given a random series of residuals.

Thus non-random residuals give high Q and *small p-value.* The significance level is related to the p-value by significance level (%) = 100 (1-p). A significance level greater than 99%, for example, corresponds to a p-value smaller than 0.01.

3.5.4 FORECASTING

The Box – Jenkins methodology requires that the model to be used in describing and forecasting a time series to be both stationary and invertible. Thus, in order to tentatively identify a Box –Jenkins model, we must first determine whether the time series we wish to forecast is stationary. If it is not, we must transform the time series into a series of stationary time series values through the process of differencing. A time series is said to be stationary (second order stationary) if the statistical properties such as the mean (first moment) and the variance (second moment) of the time series are essentially constant through time. From the plot of the time series values if the observed values of a time series seem to fluctuate with constant variation around a constant mean, then it is reasonable to believe that the time series is stationary, otherwise it is said to be nonstationary.

Notwithstanding, after scrutinizing the estimated time series model through all the diagnostic checks, then the model is fit for forecasting.

3.5.5 CONCLUSION

Chapter three is crafted into four main sections and specifically deals with the basic concepts on time series. Stationary and Non stationary time series , ARIMA model,(autoregressive integrated moving average), and principles of ARIMA modeling (Box –Jenkins 1976) the next chapter which is chapter four discusses the results of the study in a more detailed and concise manner.

CHAPTER FOUR

DATA ANALYSIS AND RESULTS

4.0 INTRODUCTION

Data analysis and results obtained from this thesis are displayed in this chapter. The analysis is discussed under the following headings; data collection and examination, model identification, diagnostic checking, and forecasting.

4.1.1 DATA COLLECTION AND EXAMINATION

Data for this thesis was collected from the KNUST solar energy laboratory. Daily data measured on hourly basis was converted to average monthly global solar radiation. Sample of the raw data is displayed at appendix one while a plot of the average monthly data is shown in figure 4.1

A time series plot of the global solar radiation of KNUST (1995-2003)

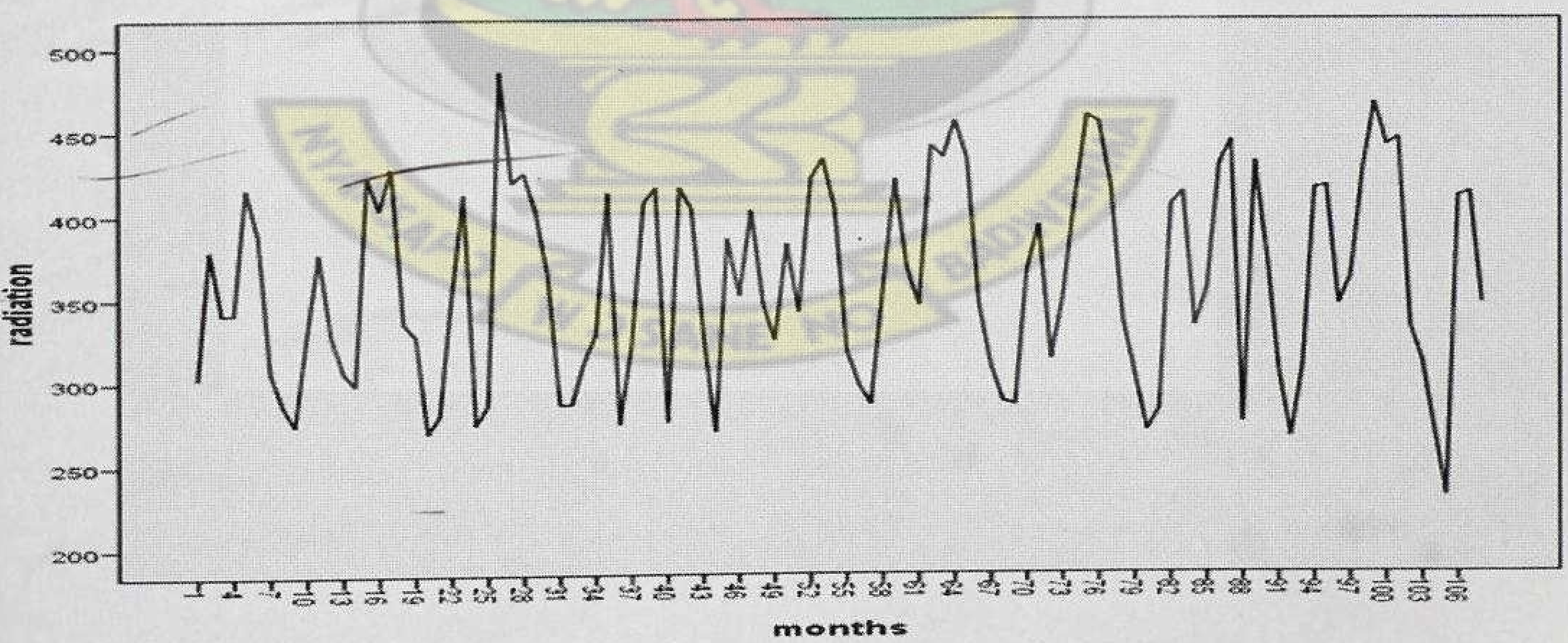


Figure 4.1 global solar radiation of KNUST (1995-2003)

4.1.2 TEST FOR STATIONARITY

We begin the analysis by testing for stationarity of the series using Kwiatkowski etal (1992)test abbreviated KPSS test. The test assesses the null hypothesis that a univariate time series y is trend stationary against the alternative hypothesis that a univariate time series y is trend non-stationary.

Table 4.1 showing kpss test for stationary

KPSS level	Truncation lag parameter	p-value
0.1599	2	0.10

From table 4.1 above since the p – value is greater than the significance level of 0.05; we fail to reject the null hypothesis which says that the data is stationaryand conclude that the data is stationary.

4.1.3 TEST FOR SEASONALITY.

The presence of an annual seasonal component in a data can be seenfrom a close examination of the ACF and the PACF plotof a time series .We therefore proceed to find the ACF and the PACF of the time series.

A plot showing the ACF against lag number of the solar radiation

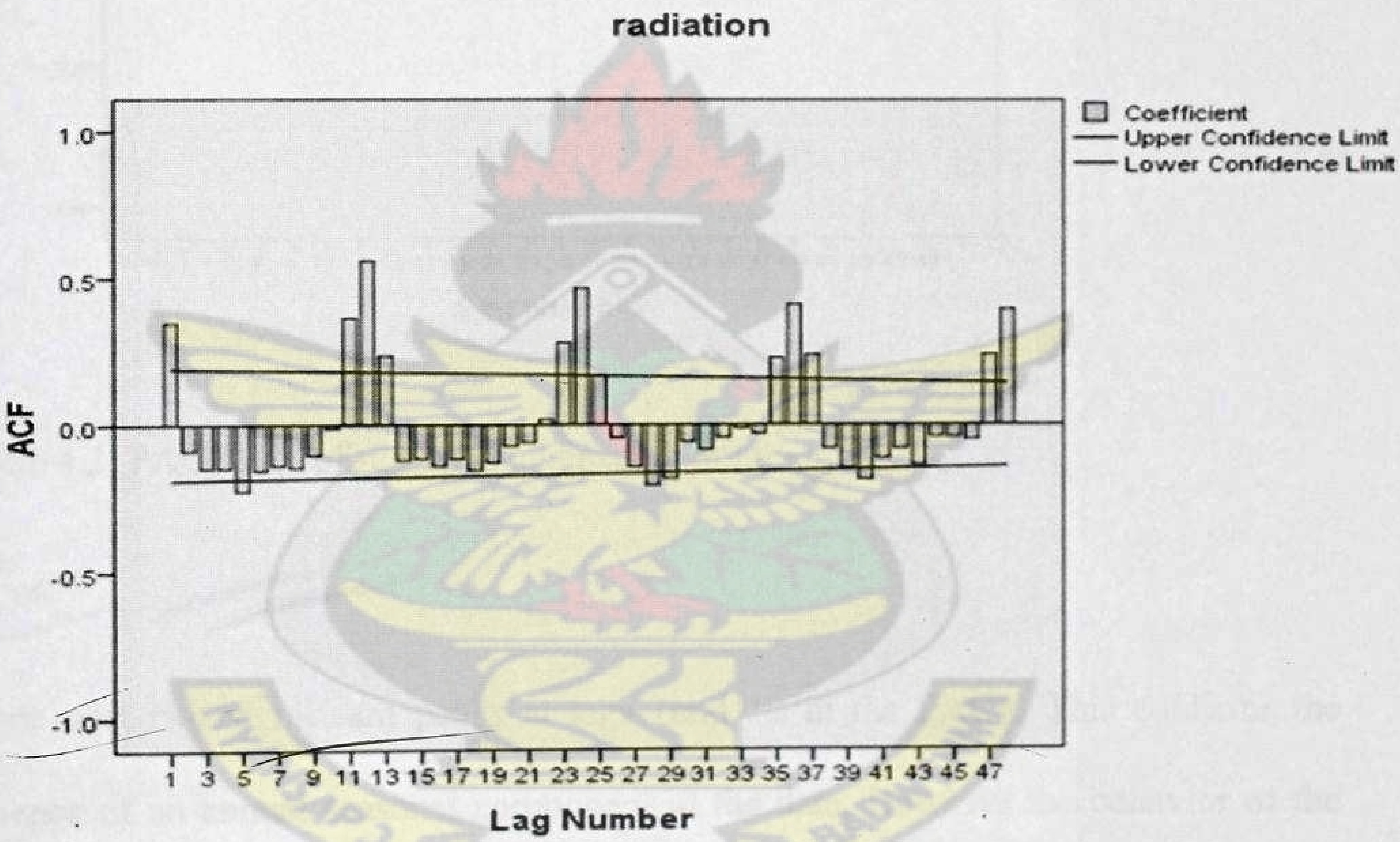


Figure 4.2: ACF plot of radiation against number of lag

The ACF showed in figure 4.3 shows significant peaks at lags of multiples of 12. This suggests the presence of an annual seasonal component in the data. However examination of the PACF will allow a more definitive conclusion.

A plot showing thePACF against lag number for solar radiation

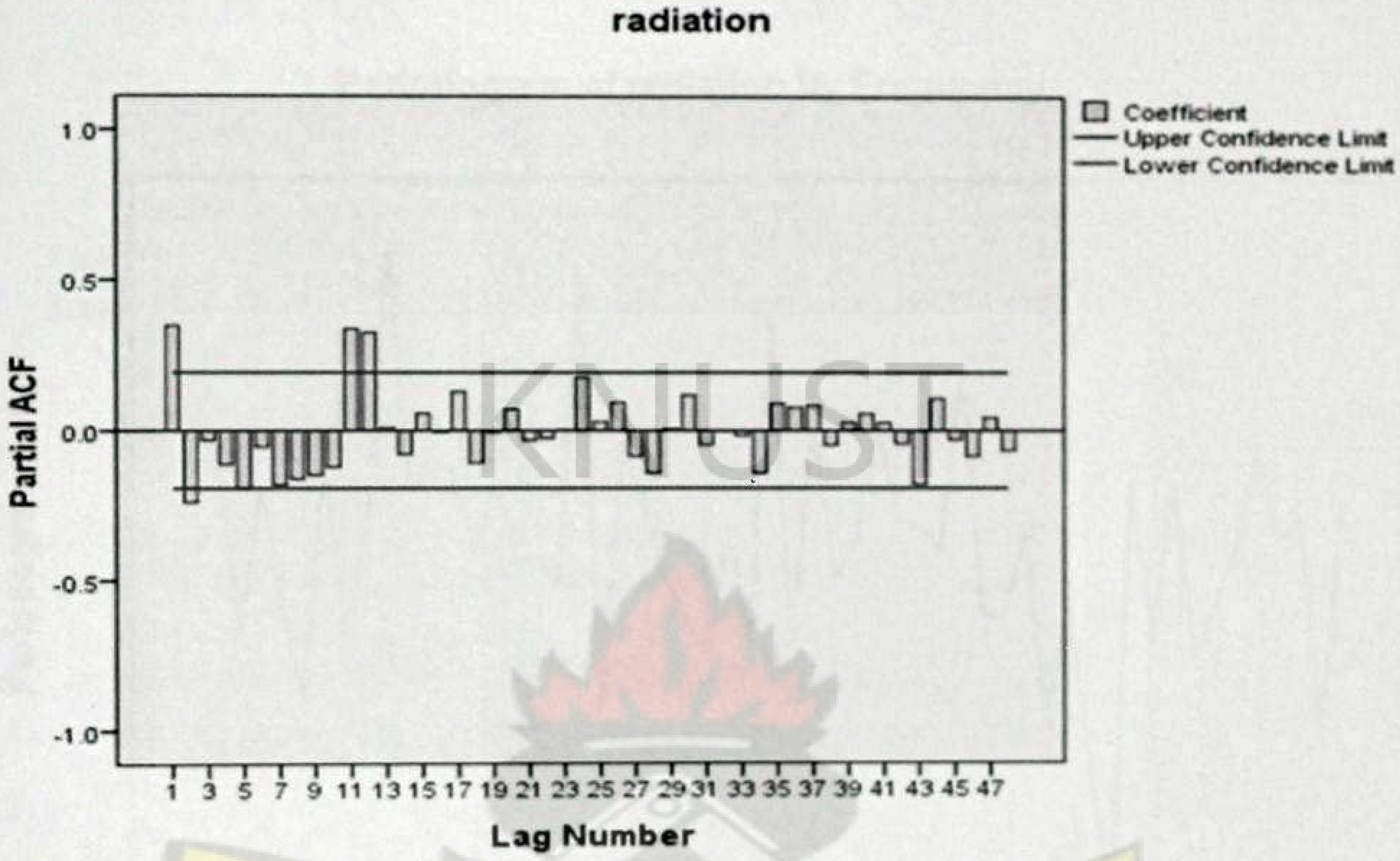


Figure 4.3: PACF plot of radiation against lag number

There exist two significant peaks at lags 1and 12 in the PACF. This confirms the presence of an annual seasonal component in the data. However the behavior of the subsequent lags indicates that the process is not purely seasonal.

4.1.4 SPECTRAL PLOTS

Spectral plot of periodogram against frequency

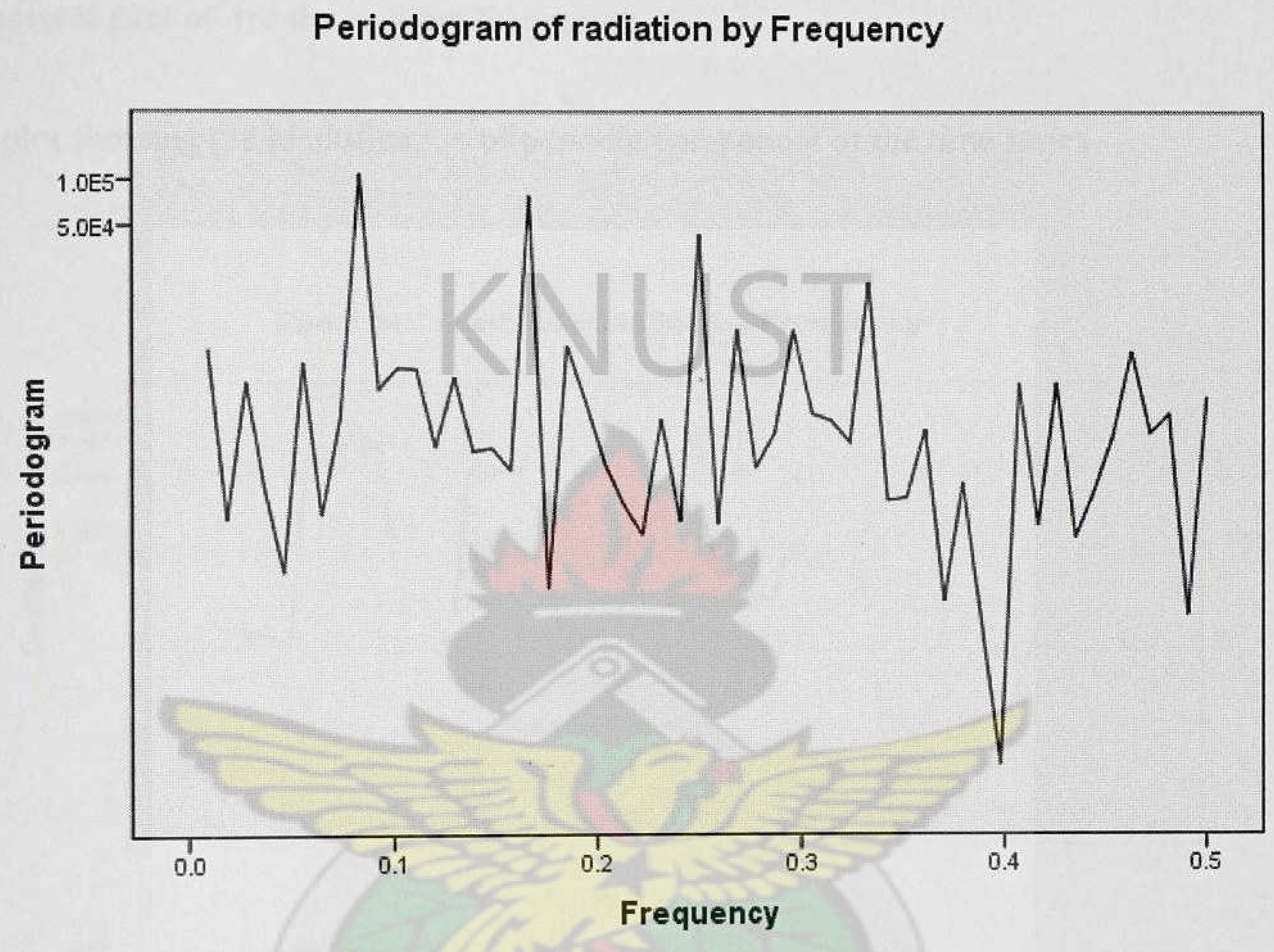


Figure 4.4: The periodogram of radiation seriesby frequency

The periodogram shows a sequence of peaks that stands out from the background noise with the frequency peak at a frequency of just less than 0.1.Each of the data points in the time series represents a month, so an annual periodicity corresponds to a period of 12.

Because period and frequency are reciprocals of each other, a period of 12 corresponds to a frequency of 1/12 (0.083)

So an annual component implies a peak in the periodogram at 0.083. And this is precisely what we expect to find if there is an annual periodic component.

Spectral plot of density against frequency

A plot showing the Identification of periodic component of the time series

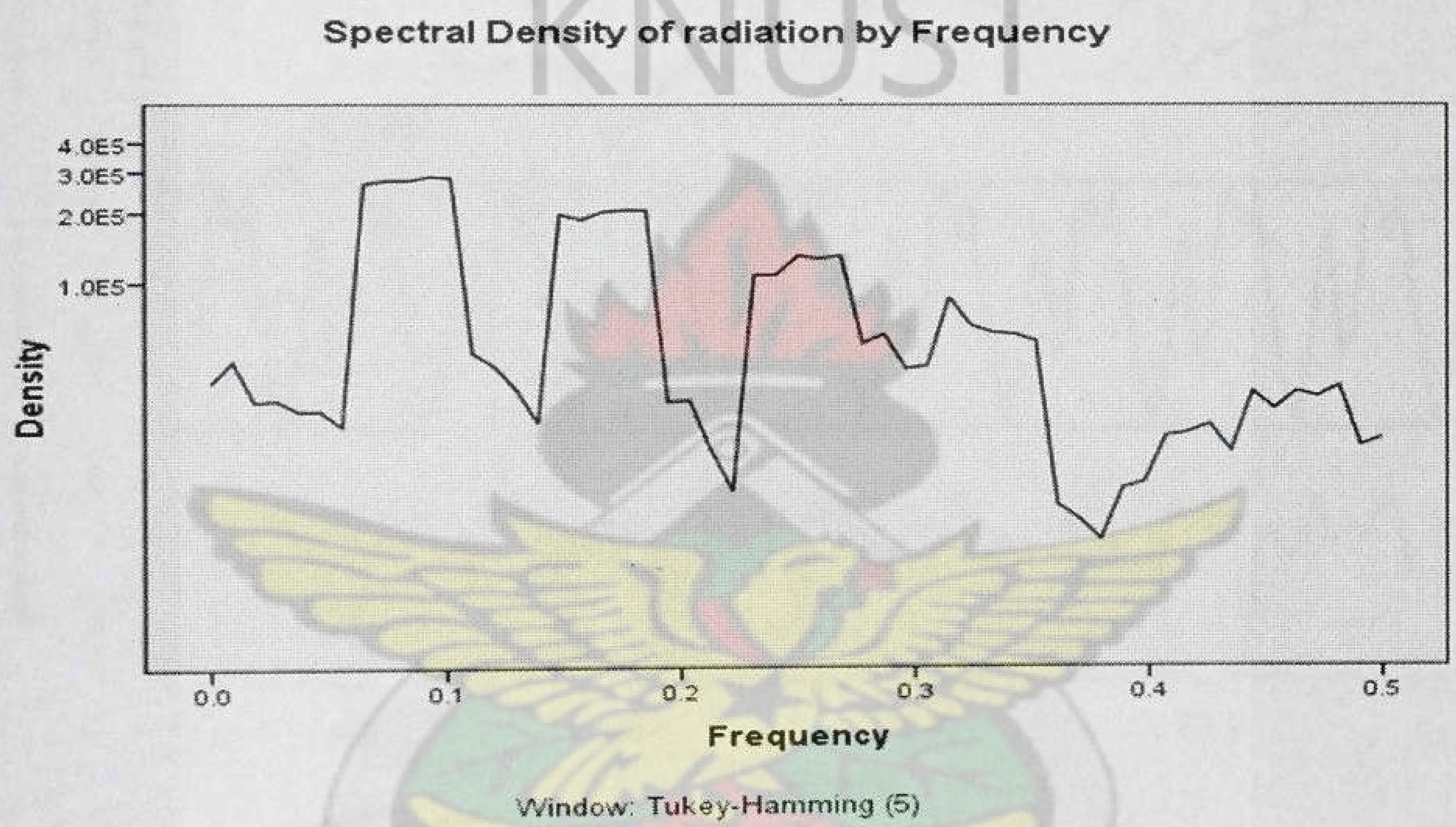


Figure 4.5: spectral ~~density of~~ radiation by frequency

Spectral density function is a smoothed version of the periodogram.Smoothing provides a means of eliminating the background noise from the periodogram allowing the underlying structure to be more clearly isolated. There are significant peaks showing at frequencies which are multiples of 0.083

We can now conclude that the data contains a periodic component with a period of 12 months.

4.1.5 DECOMPOSITION OF THE ADDITIVE TIME SERIES OF RADIATION

The figure below shows the various components of the data.

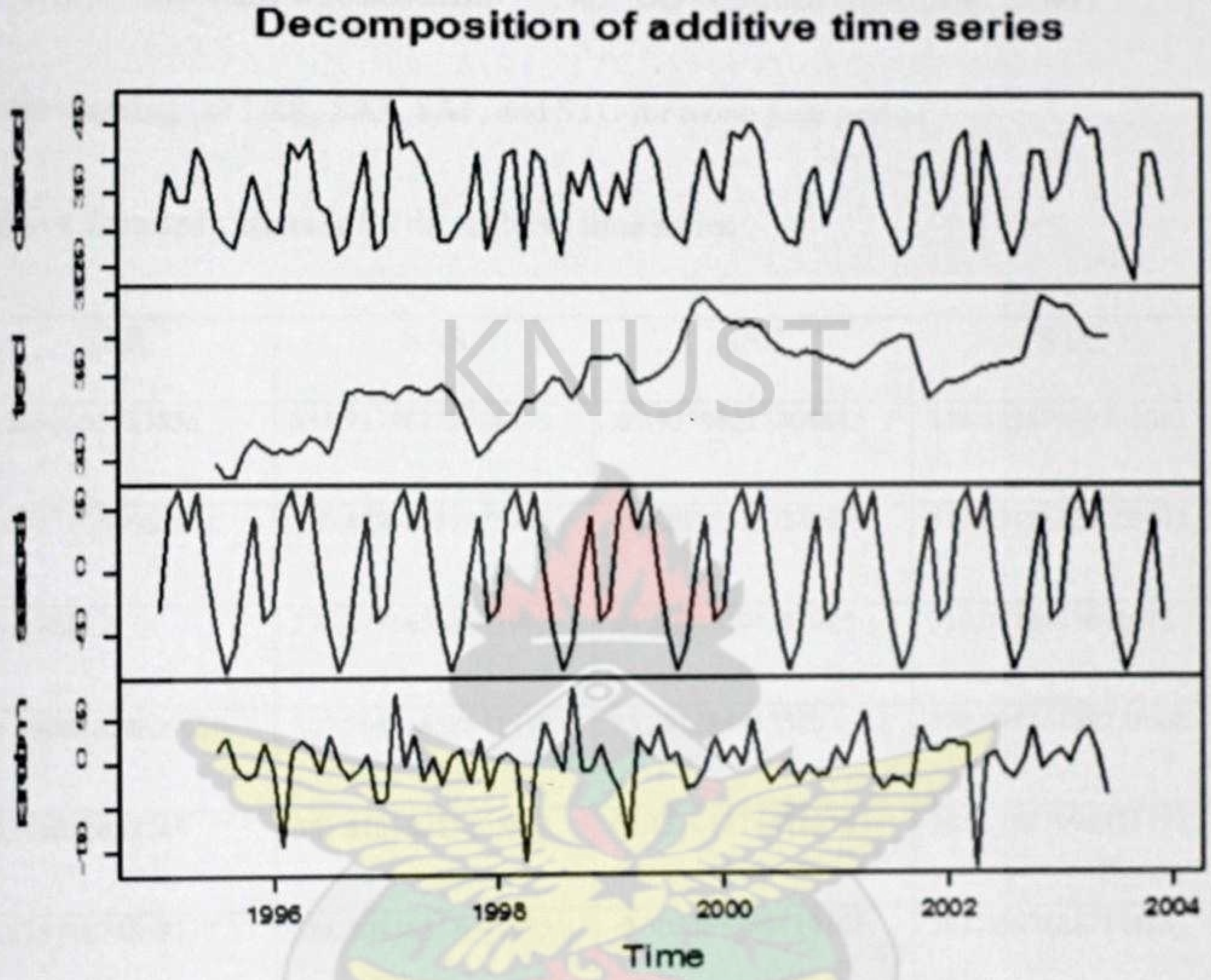


Figure 4.6: decomposition of radiation series

Decomposition of the additive time series showing from the top are the observed or original component,the trend component,the seasonal or periodic component and the random or irregular component of the series which is seen at the bottom.

We now proceed to set an annual periodicity to the series. By this the periodicity is set to 12 and this creates a set of new data variables that are designed to work with trend procedures

The set of data variables created are;

1. ERR-the residual component
2. STC-smoothed trend cycle
3. SAS-seasonally adjusted series
4. SAF-seasonal adjustment factors

Table showing the ERR, SAS, SAF, and STC for a one year period

Table 4.2created variables for the additive time series

ERR	SAS	SAF	STC
17.3993055555556	331.93798225308643	-29.93798225308642	314.53867669753083
17.44791666666663	330.47964891975306	48.52035108024692	313.03173225308643
-33.34027	276.67756558641975	64.32243441358025	310.0178433641975
-14.39004629629630	307.70881558641975	33.29118441358026	322.09886188271605
9.6349022633745	351.81298225308643	64.18701774691358	342.17807998971193
26.83577674897117	388.30256558641975	-0.302565586419761	361.4667888374486
-6.606867283950635	355.46344521604937	-49.46344521604938	362.0703125
11.944573045267532	366.03173225308643	-80.03173225308642	354.0871592078189
-9.035236625514415	331.76089891975306	-57.76089891975308	340.7961355452675
-11.81018518518516	325.93798225308643	2.0620177469135883	337.7481674382716
-6.612268518518476	332.4796489197531	44.52035108024691	339.0919174382716
32.46527	367.40673225308643	-39.40673225308642	334.94145447530866

SAF- signifies the effect of each period on the level of the series.

SAS-; signifies the original series with seasonal variations removed.

STC-; signifies a smoothed version of the seasonally adjusted series that shows both trend and cyclic component

ERR-; signifies values that remain after the seasonal, trend and cycle components have been removed from the series.

NB; The SAS is chosen to be the most appropriate series since it represents the original series with the seasonal variation removed .We thus proceed with a plot of the SAS.

A plot showing seasonal adjusted series for radiation from 1995 to 2003

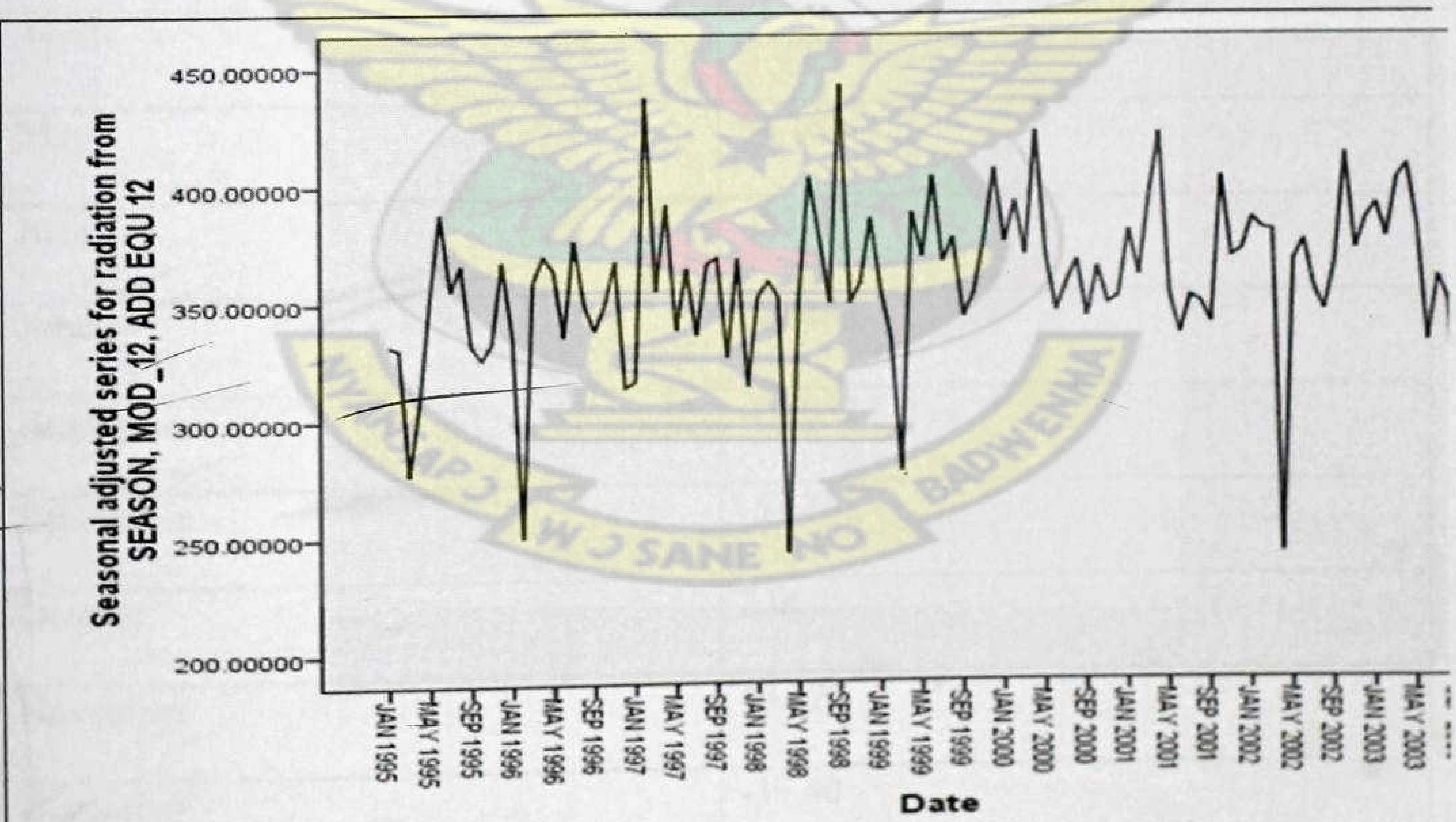


Figure 4.7: seasonally adjusted series of the radiation data.

This is the plot of the original series with the seasonal variation removed. The remaining plot reveals the trend component of the series. Figure 4.7 therefore reveals an upward trend with a number of peaks evident at random intervals, showing no evidence of an annual pattern.

4.1.6 SEASONAL FACTORS

The seasonal factors for each season or month of the year are displayed below.

Table 4.3 seasonal factors for global solar radiation of KNUST

Months	**Seasonal factor**
January	-29.938
February	48.520
March	64.322
April	33.291
May	64.187
June	-0.303
July	-49.463
August	-80.032
September	-57.761
October	2.062
November	44.520
December	-39.40

The estimated seasonal factors are given for the months January to December. They are the same for each year. The largest seasonal factor is seen in March (about

64.322), while the lowest is seen in August (about -80.032), indicating a peak in global radiation in March and a trough in August each year

4.2 MODEL IDENTIFICATION

In an attempt to identify which model is best for the data we plot the auto correlation function ACF and the partial autocorrelation function PACF as shown in the figure below.

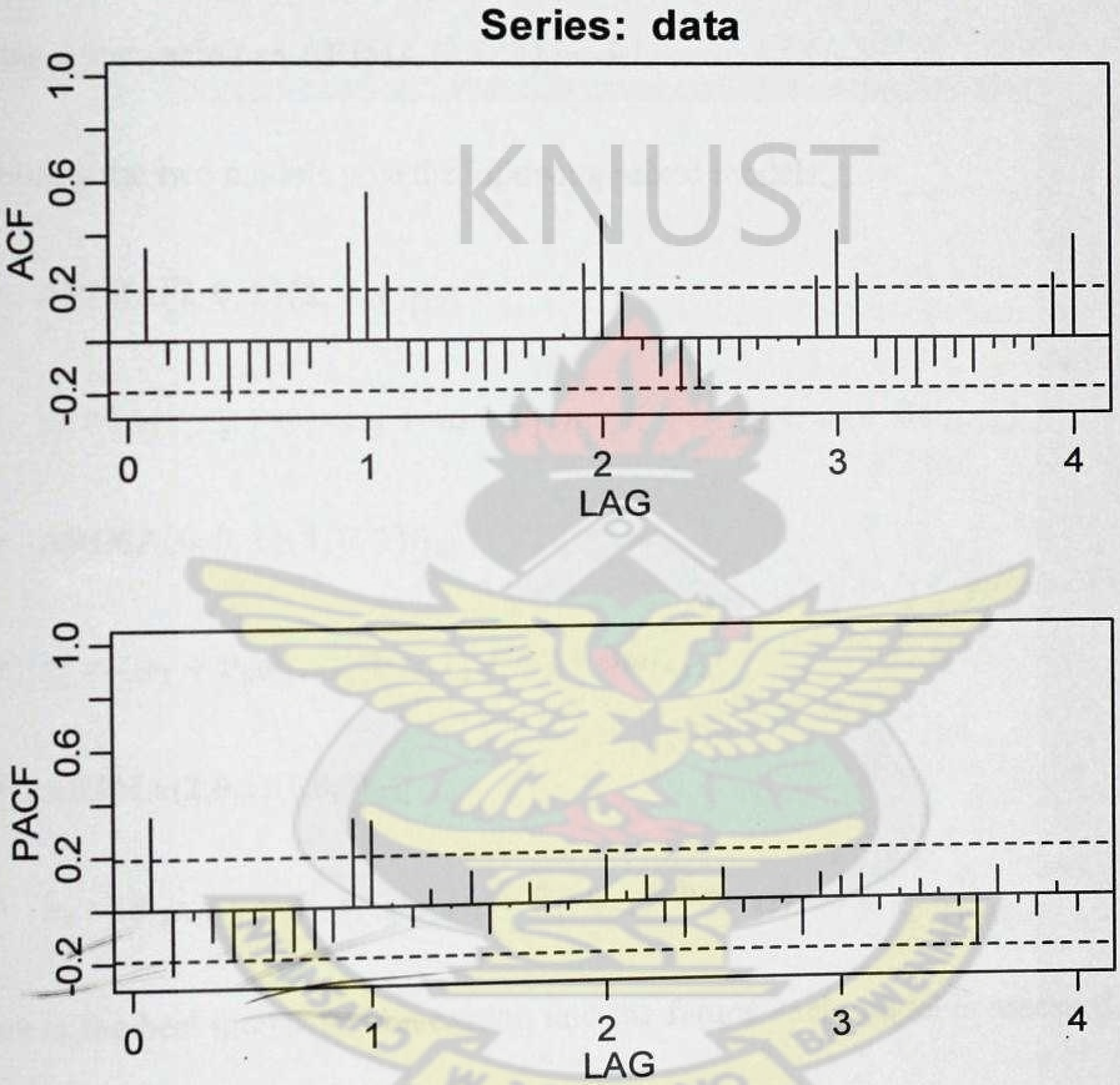


Figure 4.8: ACF and PACF of the global radiation data from 1995 – 2003

The figure above shows both the autocorrelation function and the partial autocorrelation function of global radiation at various lags.

Using the behavior of the causal and invertible pure seasonal ARIMA models, thus observing the seasonal lags, both the ACF and the PACF seem to tail off after lag k's where k=1,2,3 . . . suggesting an ARIMA (1, 0, 1)$_{[12]}$ model.

Again considering the behavior of the causal and invertible ARIMA models, thus observing the non-seasonal lags the ACF cuts off after lag 1 while the PACF cuts off after lag 2 suggesting an ARIMA (2, 0, 1) model.

Combining the two models give the following mixed models;

- $\mathbf{ARIMA(2,0,1)(1,0,1)_{[12]}}$
- $x_t = (\phi_1 x_{t-1} + \phi_2 x_{t-2} + \omega_t + \theta_1 \omega_{t-1})(\Phi x_{t-12} + \omega_t + \Theta \omega_{t-12})$
- $\mathbf{ARIMA(0,0,1)(1,0,1)_{[12]}}$
- $x_t = (\omega_t + \theta_1 \omega_{t-1})(\Phi x_{t-12} + \omega_t + \Theta \omega_{t-12})$
- **ARIMA(2,0,1)(1,0,0)$_{[12]}$**
- $x_t = (\phi_1 x_{t-1} + \phi_2 x_{t-2} + \omega_t + \theta_1 \omega_{t-1})(\Phi x_{t-12})$

To select the best model for forecasting into the future, each model is assessed based on its $\boldsymbol{AIC}, \boldsymbol{BIC}$ and$\boldsymbol{AIC_c}$..

4.2.1 MODEL SELECTION

Table 4.4 selection of best model based on Information criterion (AIC, AICc, BIC)

MODEL	AIC	AICc	BIC	Variance σ^2
ARIMA(2,0,1)(1,0,1)[12]	1128.65	1129.77	1147.43	1533
ARIMA(0,0,1)(1,0,1)[12]	1124.86	1125.44	1138.27	1525
ARIMA(2,0,1)(1,0,0)[12]	1147.7	1148.53	1163.79	2066

The table above shows the Akaike Information criterion (AIC), the corrected Akaike Information criterion (AICc) and the Bayesian Information criterion (BIC) respectively.

The information criterion indicate that the model with the least value is selected among the other proposed models

From the tableabovethe best model among the suggested models is ARIMA (0,0,1)(1,0,1)[12] since it has the smallest value for the AIC, AICc and BIC. It also has the smallest variance

4.2.2 MODEL FITTING

The co-efficient, the standard error (s.e) and the t-values for the various ARIMA orders are displayed in the table below;

Table 4.5 parameter estimates of $\boldsymbol{ARIMA(0,0,1)(1,0,1)_{[12]}}$

	ma1	sar1	sma1	x-mean
Coefficient	0.1487	0.9835	-0.7886	358.3099
s.e	0.1009	0.0203	0.1262	15.2334
t-values	1.4737	48.4483	6.2488	

The common rule or assumption is that a t-statistics with an absolute value greater than 1.25 for lags 1 through3 or greater than 2 for lags 4 and beyond indicates a coefficient statistically significant.

Based on the t-value test all the coefficients aresaid to be statistically significant since their t-values are all greater than 1.25.

4.3 MODEL DIAGNOSTICS

Diagnostics of $\boldsymbol{ARIMA(0,0,1)(1,0,1)_{[12]}}$

The figure shows the following; (four in one plot labeled a, b, c, and d)

a. standardized residuals

b. ACF of residuals

c. Normal Q-Q plot of standardized residuals

d. P-values for Ljung-Box statistics

444ofej

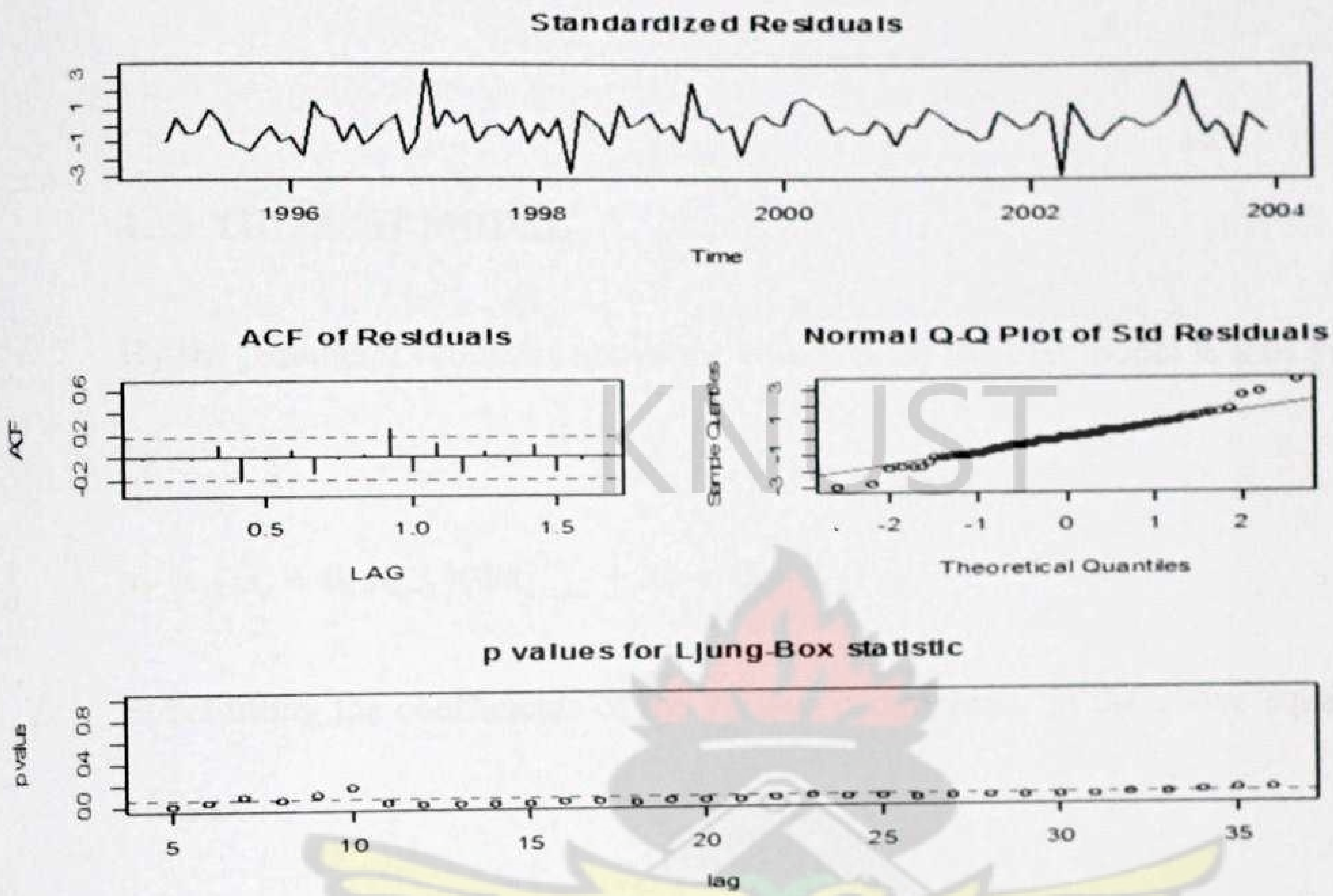


Figure 4.9: DIAGNOSTICS OF **ARIMA(0, 0, 1)(1, 0, 1)$_{[12]}$**

a. The standardized residuals plot in the figure above supports the model, since there is no evidence of trend, thus the standardized plot exhibits no obvious pattern.

b. From the plot of the ACF of the residuals; the lags show no statistically significant evidence of non – zero autocorrelation in the residuals.

c. With the normal q- q plot of the standardized residual, most of the residuals seem to follow the line of best fit fairly closely except for some few residuals deviating from the normality. Since most of the residuals are located on the straight line, we conclude that the normality assumption has also been satisfied.

d. The plot of the Ljung – Box statistic shows that most of the Ljung – Box p – values are all less than 0.05, thus the Ljung – Box statistic is significant at any positive lag.

4.3.1 THE BEST MODEL

By the parameter estimates above the equation for the best model is thus given by

$$x_t = (\omega_t + \theta_1\omega_{t-1})(\Phi x_{t-12} + \omega_t + \Theta\omega_{t-12})$$

Substituting the coefficients of the various parameters in the above equation yields

$$x_t = (\omega_t + 0.1487\omega_{t-1})\,(0.9835x_{t-12} + \omega_t - 0.7886\omega_{t-12})$$

4.4 FORECASTING

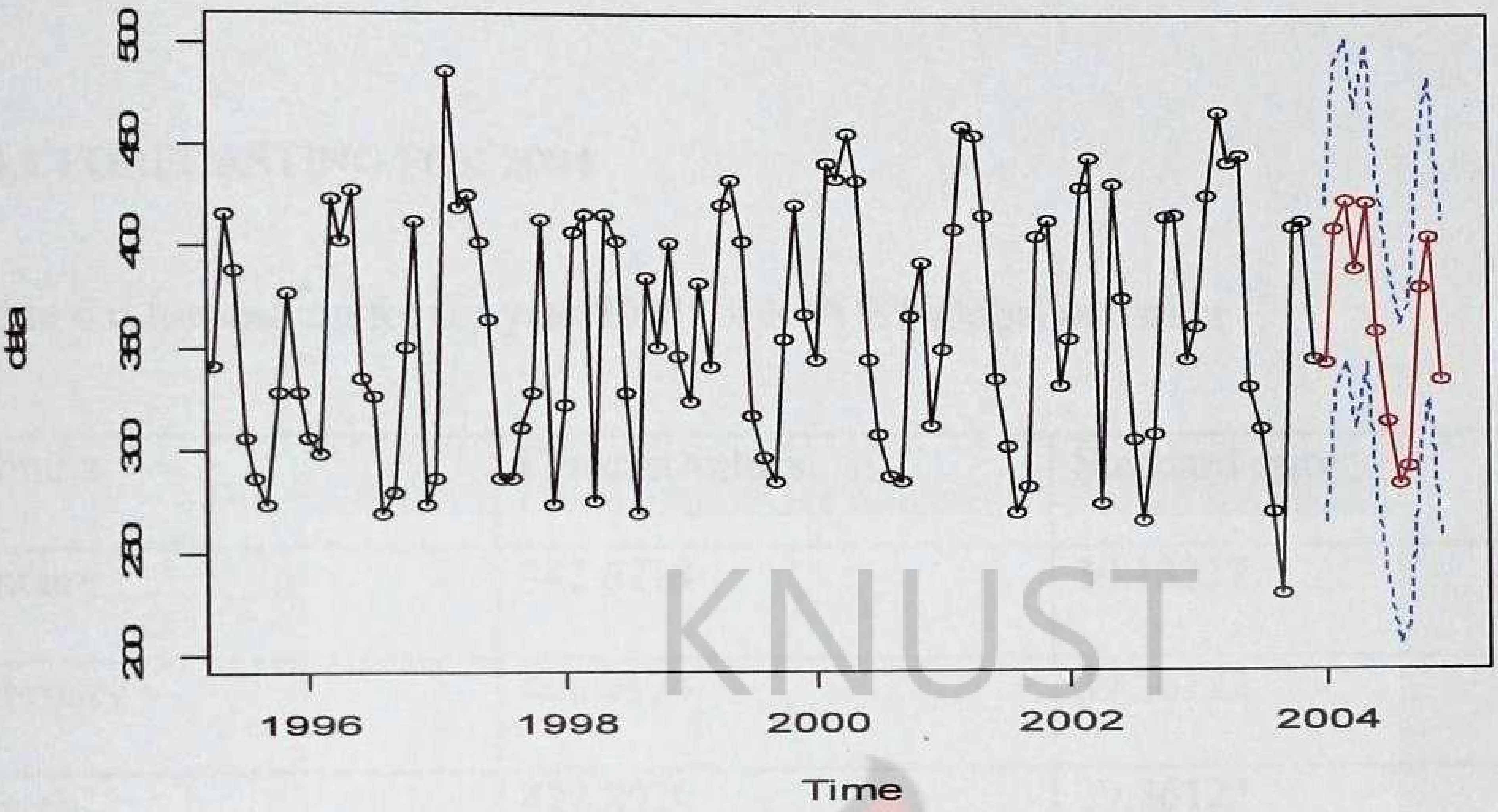


Figure 4.10a: radiation forecasts with confidence intervals for 2004

Figure above gives the visual representation of the original global radiation data, its forecasts and confidence interval.

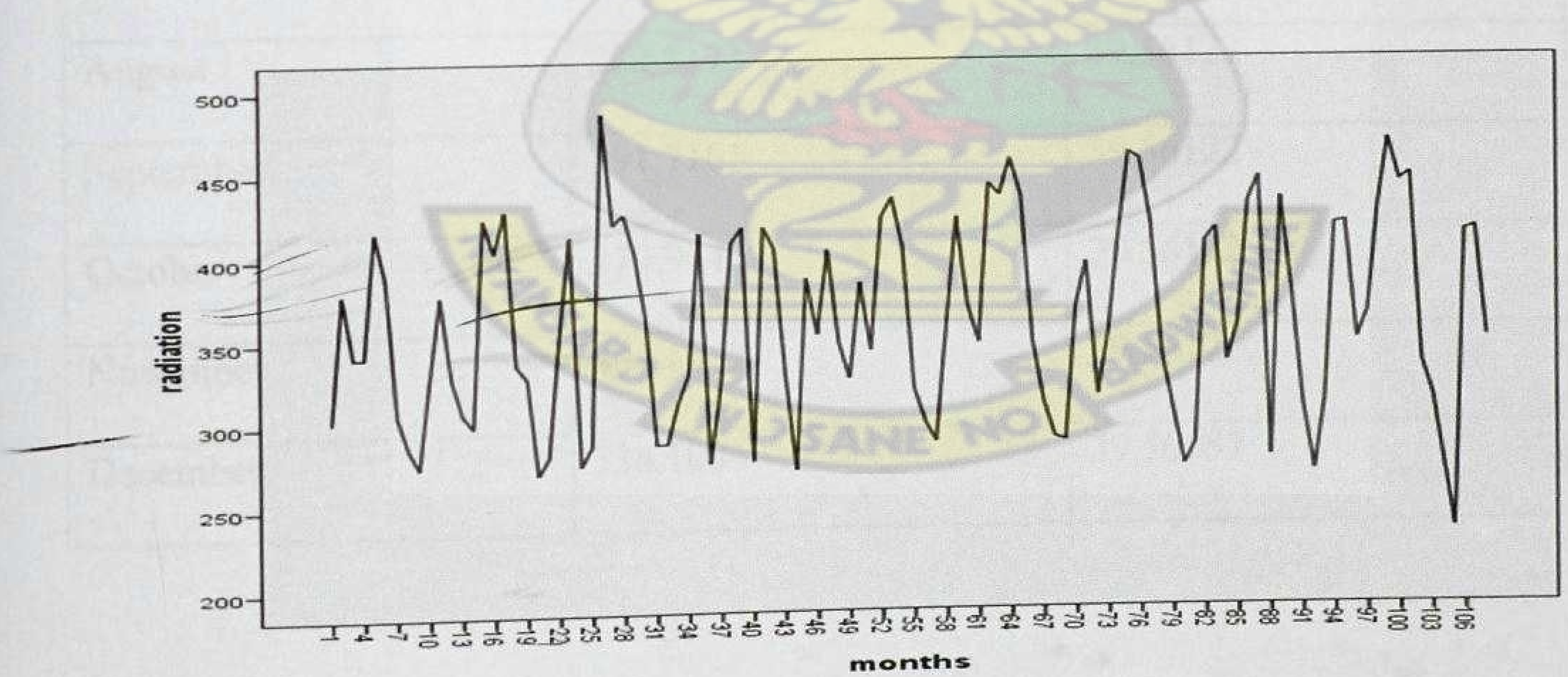


Figure 4.10b showing radiation plot with existing 2004 data.

Based on the supporting model statistics and forecasting criterion, it is proposed that the best model among the three suggested models is ARIMA (0,0,1)(1,0,1)[12]

4.4.1 FORECASTING FOR 2004

Table 4.6 forecasting for the year 2004 for KNUST global radiation

Months	Forecast values	Standard error
January	342.6218	39.13217
February	408.7124	39.56122
March	422.8910	39.56122
April	388.7332	39.56122
May	421.7421	39.56122
June	358.6227	39.56122
July	313.6345	39.56122
August	282.8173	39.56122
September	291.7101	39.56122
October	379.9046	39.56122
November	404.1462	39.56122
December	334.1084	39.56057

4.4.2 EVALUATIONOF FORECAST PERFORMANCE

The first test for forecasting models were developed in 1939 by Tinbergen in response to Keynes who stated that theories must be confirmed if the data and statistical methods are employed correctly. The standard approach of forecast accuracy analysis is to investigate the bias of forecasts, their efficiency in terms of incorporating all available information and performance compared with other forecast of the same indicator. Test of forecast efficiency, determines whether the forecast utilizes all the available information at a given point in time. It is further to check whether the forecast efficiency is related to the question of the impact of potential determinants of radiation on the forecast performance. To assess the out-of-sample forecasting ability of the model it is advisable to retain some observations at the end of the sample period which are not used to estimate the model. Consider for example the radiation data from the period of January 1995 to December 2003, which was used to model the radiation and subsequently provided twelve months ahead forecast from January to December 2004 as shown below.

Table 4.7 forecasting performance evaluation

PERIOD	FORECAST VALUES	ACTUAL VALUES	ERROR	% ERROR	LOWER	UPPER
109	343	335	-8	2.4	241.844	458.464
110	409	414	-5	1.2	240.377	476.641
111	423	450	27	6	240.377	476.641
112	389	414	25	6	240.377	476.641
113	422	383	-39	10.2	240.377	476.641
114	359	341	-18	5.3	240.377	476.641
115	314	273	-41	15	240.377	476.641
116	283	242	-42	17.4	240.377	476.641
117	292	355	63	17.7	240.377	476.641
118	380	403	23	5.7	240.377	476.641
119	404	402	-2	0.005	240.377	476.641
120	334	396	62	15.7	240.377	476.641

The strength or power of a forecast is determined by the calculated forecast error. The error of a forecast is calculated by the difference between the actual values and the forecast values. When the forecast value is higher than the actual value the error is negative and we say the forecast is high. Similarly we say the forecast is low if the forecast value is of a lower value than the actual value.

4.5 SUMMARY OF FINDINGS

The thesis was aimed at selecting the best model among various ARIMA estimated models which has the highest forecasting power. We have identified a framework for ARIMA modeling which includes the following steps; data collection and examination; model identification; diagnostic checking and forecasting performance evaluation. We adopted the traditional Box-Jenkins approach of forecasting known as ARIMA modeling, in which a time series is expressed in terms of past values of itself (the autoregressive component) plus current and lagged values of a 'white noise' error term (the moving average component). The primary purpose behind this study was to find out which ARIMA model is more accurate and appropriate for forecasting purposes in the real world situation, keeping in mind the cost of modeling building.

A general rule of thumb for univariate forecasting is to test for all the stages of the ARIMA process. ARIMA models are theoretically justified and can be surprisingly robust with respect to alternative (multivariate) modeling approaches. The study was based on KNUST average monthly global solar radiation data, which was used to estimate various possible ARIMA models and the candid model, was selected based on the BIC, AIC, and the AIC_C. The comparative performance of these ARIMA models have been checked and verified. Several ARIMA models were selected and tested and it was concluded that $ARIMA(0,0,1)(1,0,1)_{[12]}$ is the best model.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 CONCLUSION

The objective of this thesis is to model the average monthly global radiation of KNUST using time series analysis and also to predict the future global radiation of KNUST.

By the application of the Box-Jenkins methodology the analysis was successfully carried out with the following results;

1. The model developed is ARIMA$(0,0,1)(1,0,1)_{[12]}$ or

- $x_t = (\omega_t + \theta_1\omega_{t-1})(\ \Phi x_{t-12} + \omega_t + \Theta\omega_{t-12})$

Substituting the values of the estimated parameters; we have

- $x_t = (\omega_t + 0.1487\omega_{t-1})(\ 0.9835x_{t-12} + \omega_t - 0.7886\omega_{t-12})$

2. From the estimated seasonal factors shown in table 4.1, it is revealed that global solar radiation attains its peak in March while the lowest is recorded in august.

5.2 RECOMMENDATIONS

The modeled global solar radiation shall serve the following purpose for KNUST and for the matter the nation as a whole.

1. To generate global monthly average solar radiation data for KNUST.

2. for proper assessment of the solar energy resource potential of KNUST.

3. Serve as the basis for engineering and economic decisions, for example, in the construction of solar based devices such as solar panels, and solar water heaters.

4. Maximum solar energy could be stored in March where radiation is at its peak and used for a number of solar energy applications such as photovoltaic systems for electricity generation during low radiation seasons at the energy center.

5.3 LIMITATION

1. At the time of writing the thesis available data was from 1995 to 2004 only.

2. Forecasted values were already in existence.

REFERENCE

1. Akuffo et al (2003); solar and wind energy resource assessment (SWERA)
2. Almorox and Hontoria (2003) Global solar radiation estimation using sunshine duration in spain
3. Anderson B. (1976); an implementation of the Burnham-Anderson distribution-free estimator
4. Assi and Jama (2010) prediction of monthly average daily global solar radiation in Al Ain City UAE using Artificial Neural Networks
5. Bar-Sever1 and Kuang1 (2004) New Empirically Derived Solar Radiation Pressure Model for Global Positioning System Satellites
6. Box and Jenkins (1976); Box- Jenkins time series analysis.
7. Bulut and Buyukalaca (2006) Simple model for the generation of daily global solar-radiation data in Turkey; vol 84 Issue 5, May 2007, pages 477-491
8. Chatfield J. (1984); the analysis of time series
9. Cryer and chan (2008); time series analysis with application in R second Edition.
10. El-Sebaii1and Trabea 2 (2005) Estimation of Global Solar Radiationon Horizontal Surfaces Over Egypt.
11. Ertekin and Yaldiz (1999); estimation of monthly average daily global radiation on horizontal surface for Antalya (Turkey)

12. Falayi et al. (2008) Empirical models for the correlation of global solar radiation with meteorological data for Iseyin, Nigeria: International Journal of Physical Sciences Vol. 3 (9), pp. 210–216, September 2008

13. Fooladman and H. R (2012); evapotranspiration using actual and estimated sunshine hours in south of Iran; Vol. 7(7), pp. 1164-1169, 19 February, 2012 Available online at http://www.academicjournals.org/AJA

14. Isikwue et al (2012) Empirical Model for the Estimation of Global Solar Radiation in Makurdi, Nigeria

15. Judge G. G. (1985); the theory and practice of Econometrics

16. Karoro et al (2011) Predicting Global Solar Radiation Using an Artificial Neural Network Single-Parameter Model

17. Kwiatkowski et al (1992): Testing the Null Hypothesis of Stationarity against the

18. Linares-Rodríguez et al (2011)Generation of synthetic daily global solar radiation data based on ERA- Interim reanalysis and artificial neural networks :journal homepage: www.elsevier.com/locate/energy

19. Markvart etal (2005) PV system sizing using observed time series of solar radiation

20. Martı´n et al (2010) Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning Solar Energy 84 (2010) 1772–1781;

21. Mihalakakou et al (1999) the total solar radiation time series simulation in Athens, using neural networks; Theoretical and Applied Climatology. 66, 185±197 (2000)

22. Myers D. R. (2003) Solar Radiation Modeling and Measurements for Renewable Energy Applications:

23. Okundamiya and Nzeako (2011) Empirical Model for Estimating Global Solar Radiation on Horizontal Surfaces for Selected Cities in the Six Geopolitical Zones in Nigeria.: Research Journal of Applied Sciences, Engineering and Technology 2(8): 805-812, 2010:

24. Paoli1 et al (1970) Solar radiation forecasting using ad-hoc time series preprocessing and neural networks 1

25. Paulescu and Schlett (2003); Performance assessment of global solar irradiation models under Romanian climate

26. Raji et al (2012), Analysis of Global Solar Radiation Data as Time Series Data for Some Selected Cities of Western Part of Nigeria; international journal of advanced renewable energy research Vol. 1 Issue. 1, pp. 14-19, 2012

27. Remund1and Müller1 (1975) Trends in global radiation between 1950 and 2100

28. Safari and Gasore (2009) Estimation of Global Solar Radiation in Rwanda using Empirical Models. Asian Journal of Scientific Research 2(2); 68-75, (2009)

29. Saffaripour and Mehrabian, (2009); 'Numerical methods applied to global solar radiation modeling-comparison with measured data'

30. Sun and Kok (2007) a solar radiation model with a Fourier transform approach

31. Wong and Chow (2001) solar radiation model. Applied energy vol. 69 pp. 191-224

32. Yorukoglu and Celik (2005) A critical review on the estimation of daily global solar radiation from sunshine duration

33. Zaharim et al (2009) Time Series Analysis of Solar Radiation Data in the Tropics

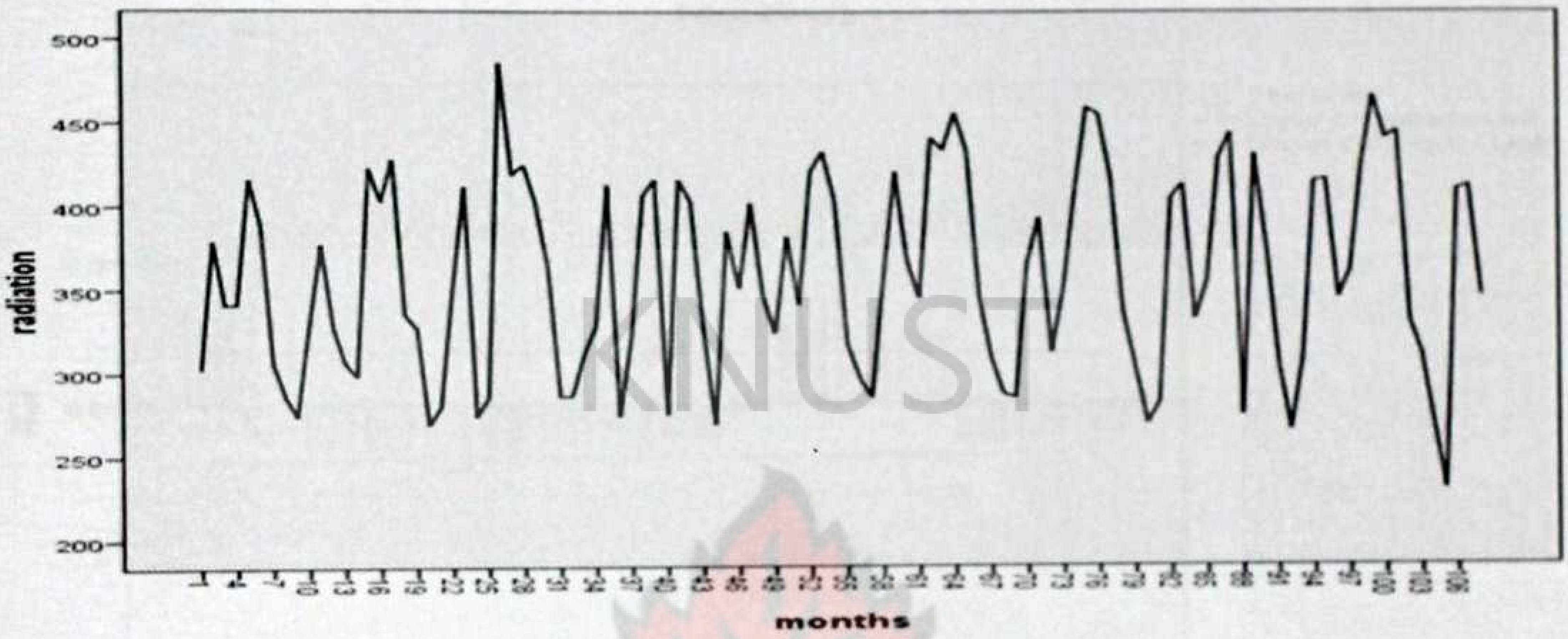
APPENDIX 1

SAMPLE OF RAW RADIATION DATA

Month	Time	Global/Ave Rad (W/m^2)	Diff/Ave Rad (W/m^2)	Temp (^{0}C)	Gmim	Press (bar)
1	100	-2.55	-1.254	59	-3.135	1
1	200	-1.94	-1.045	101	-2.717	136
1	300	-0.641	-0.418	236	-1.045	204
1	400	-0.422	-0.418	301	-0.627	306
1	500	-0.526	-0.418	401	-0.836	456
1	600	-1.098	-0.627	523	-1.673	544
1	700	2.526	14.22	700	-1.673	601
1	800	36.92	72.3	800	15.05	701
1	900	133.9	294.7	859	70	807
1	1000	250.6	435.9	944	184.2	934
1	1100	277.4	492.2	1043	222.6	1006
1	1200	458.8	603.9	1126	261.2	1103
1	1300	485	650.6	1247	314.9	1233
1	1400	540.6	627.6	1306	275.2	1400
1	1500	323.3	501.6	1418	188.4	1450
1	1600	249.6	314.4	1524	184.2	1503
1	1700	157.3	237.6	1606	87.4	1655

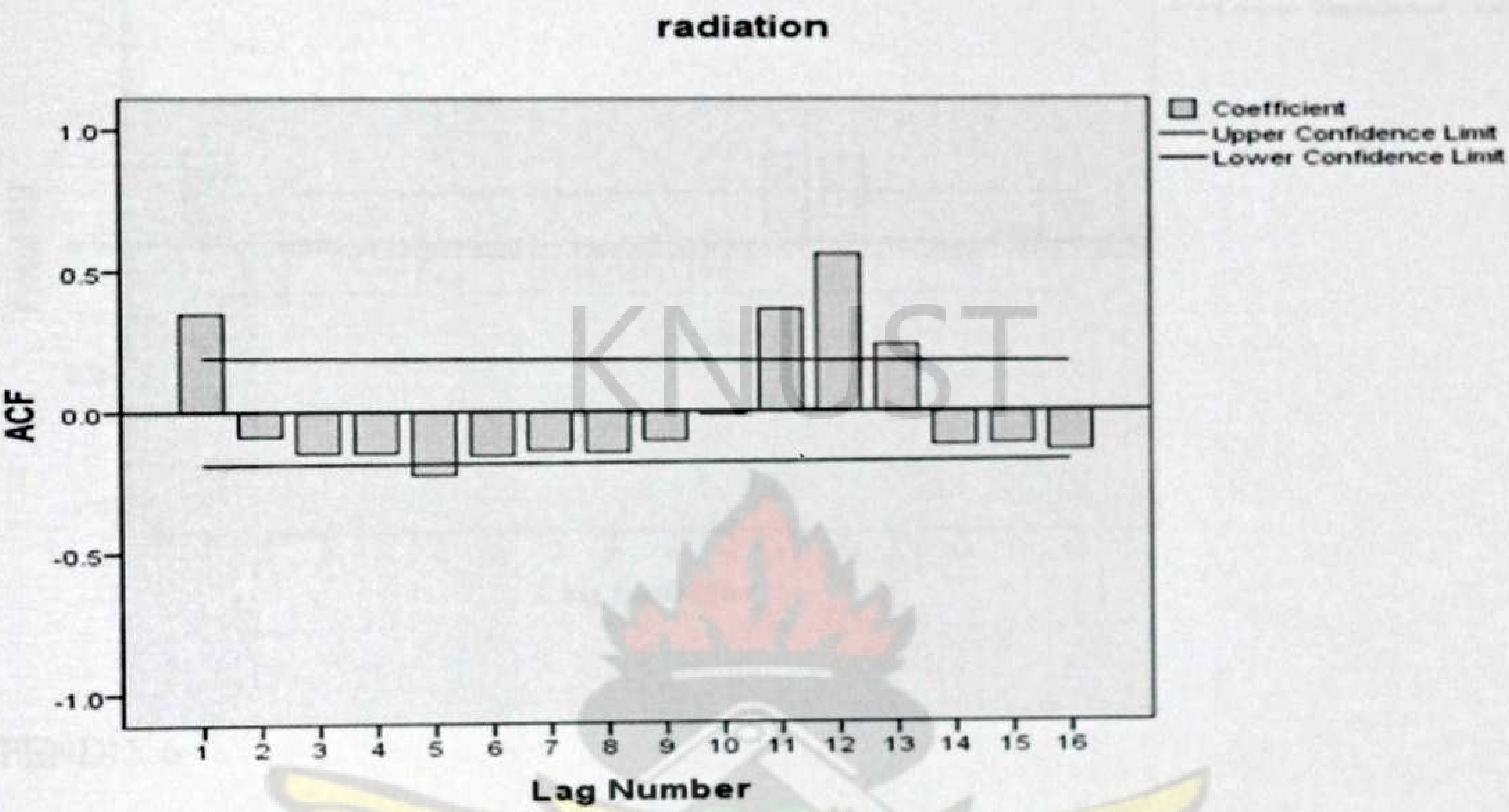
APPENDIX 2

Time series plot of global radiation



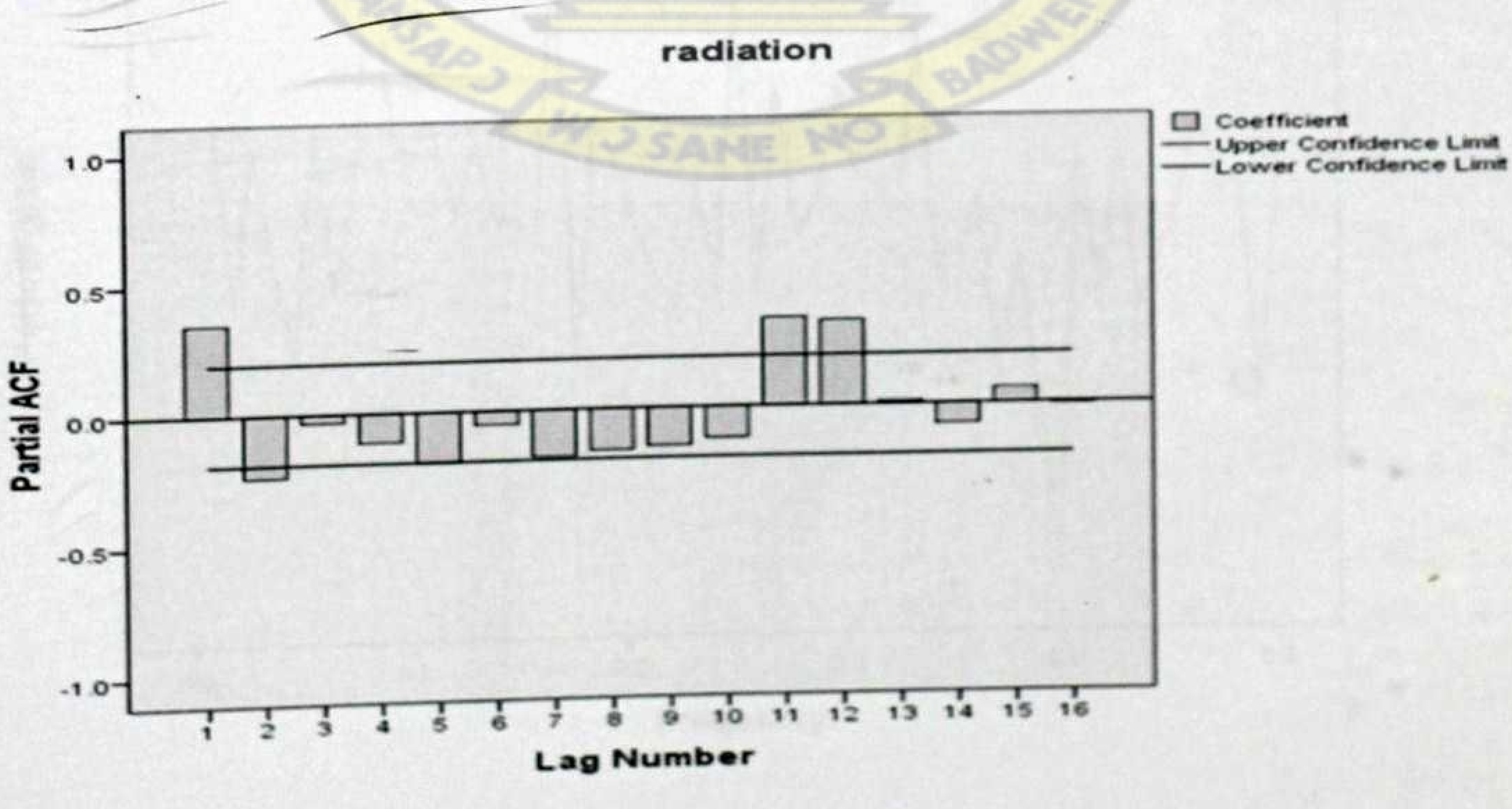
APPENDIX 3

ACF of global radiation

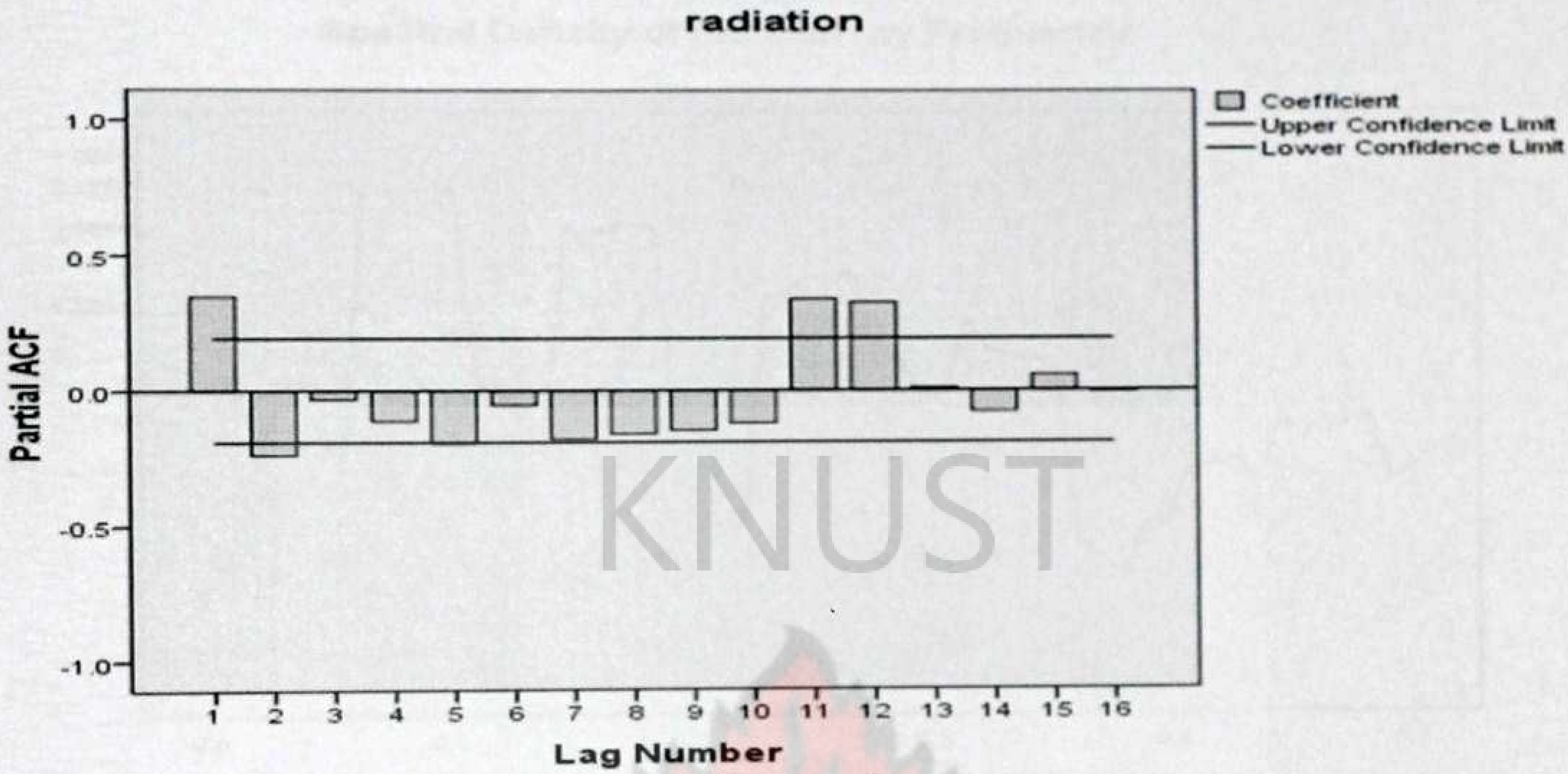


APPENDIX 4

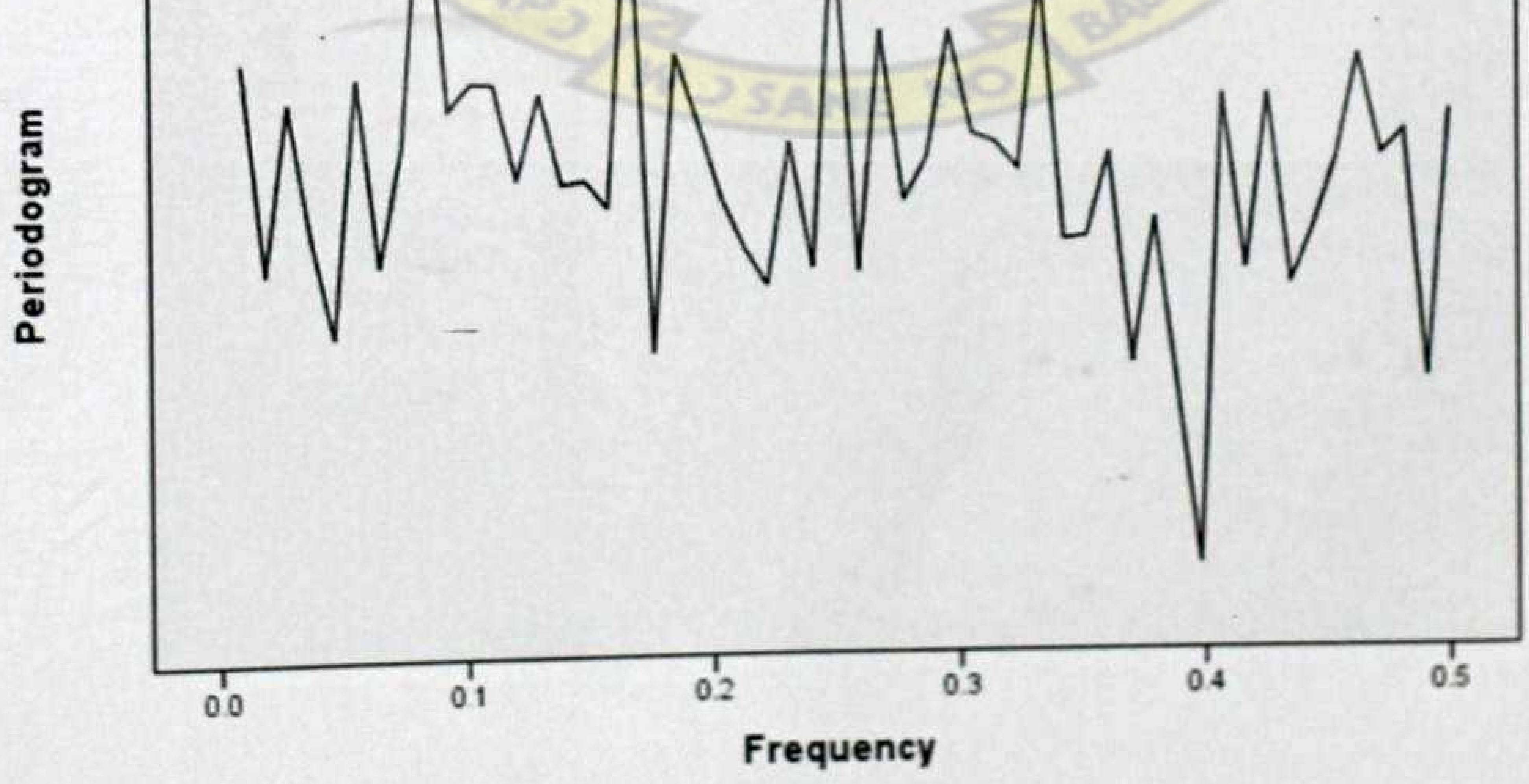
PACF of global radiation



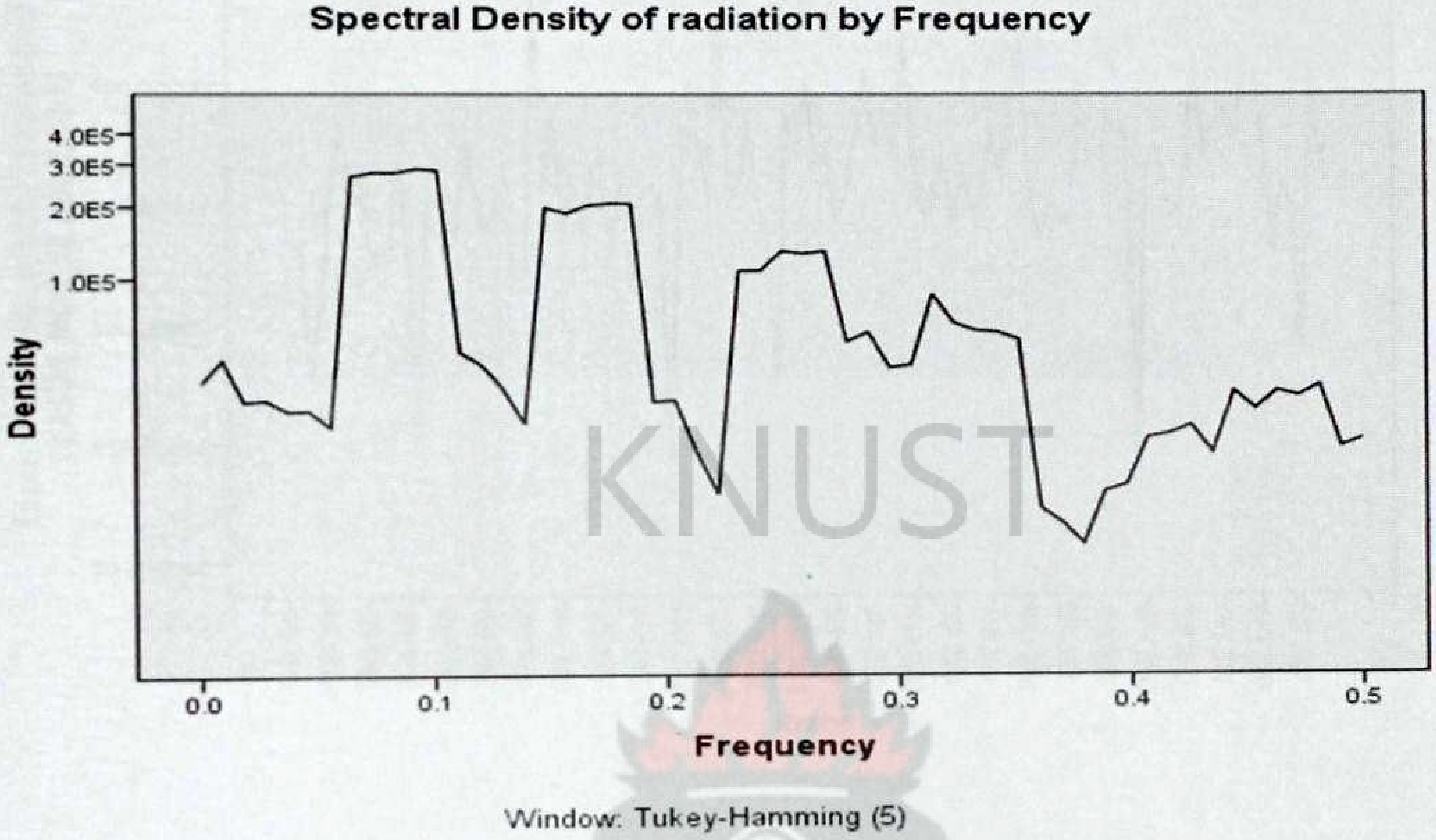
APPENDIX 5



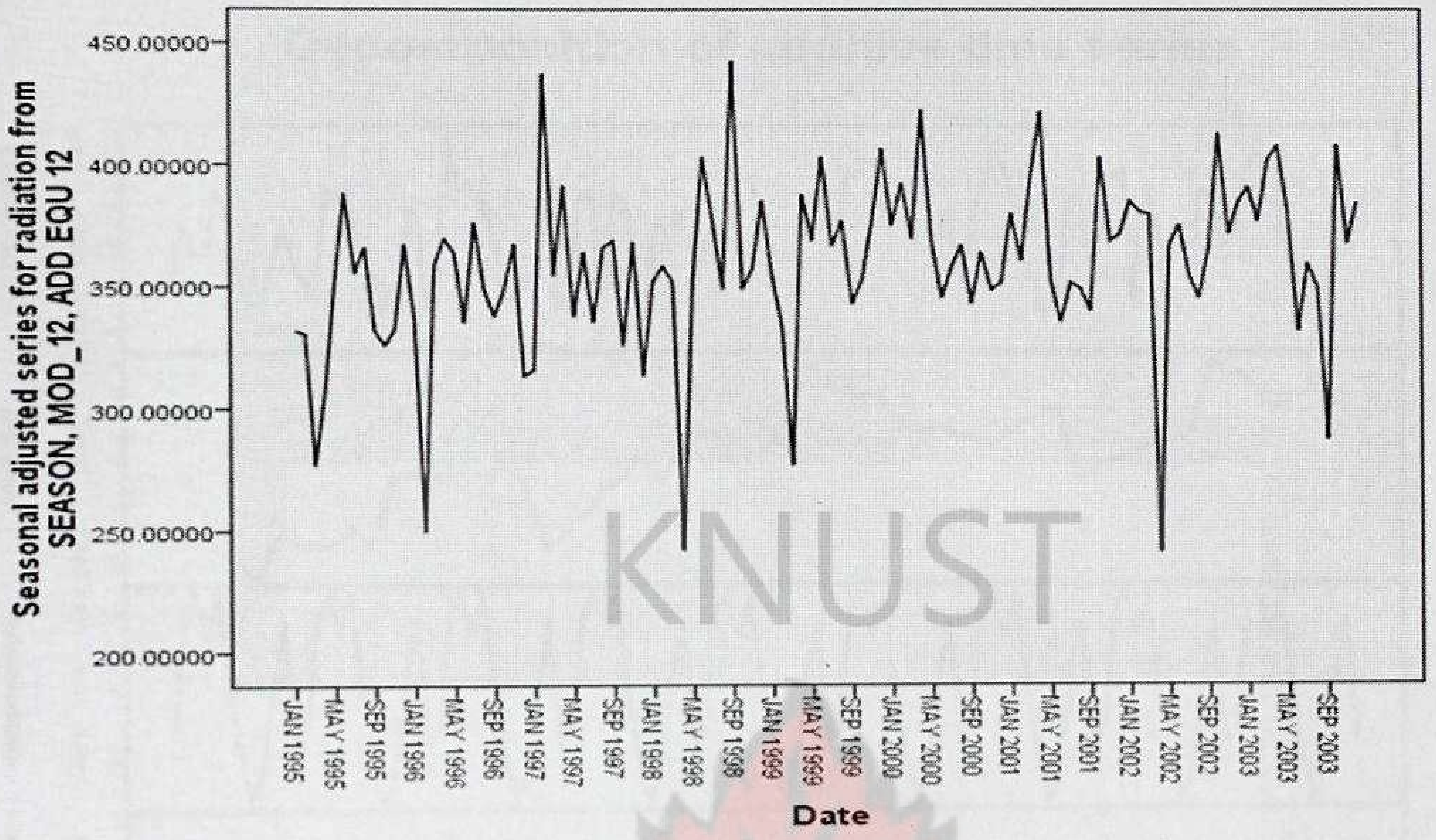
APPENDIX 6



APPENDIX 7

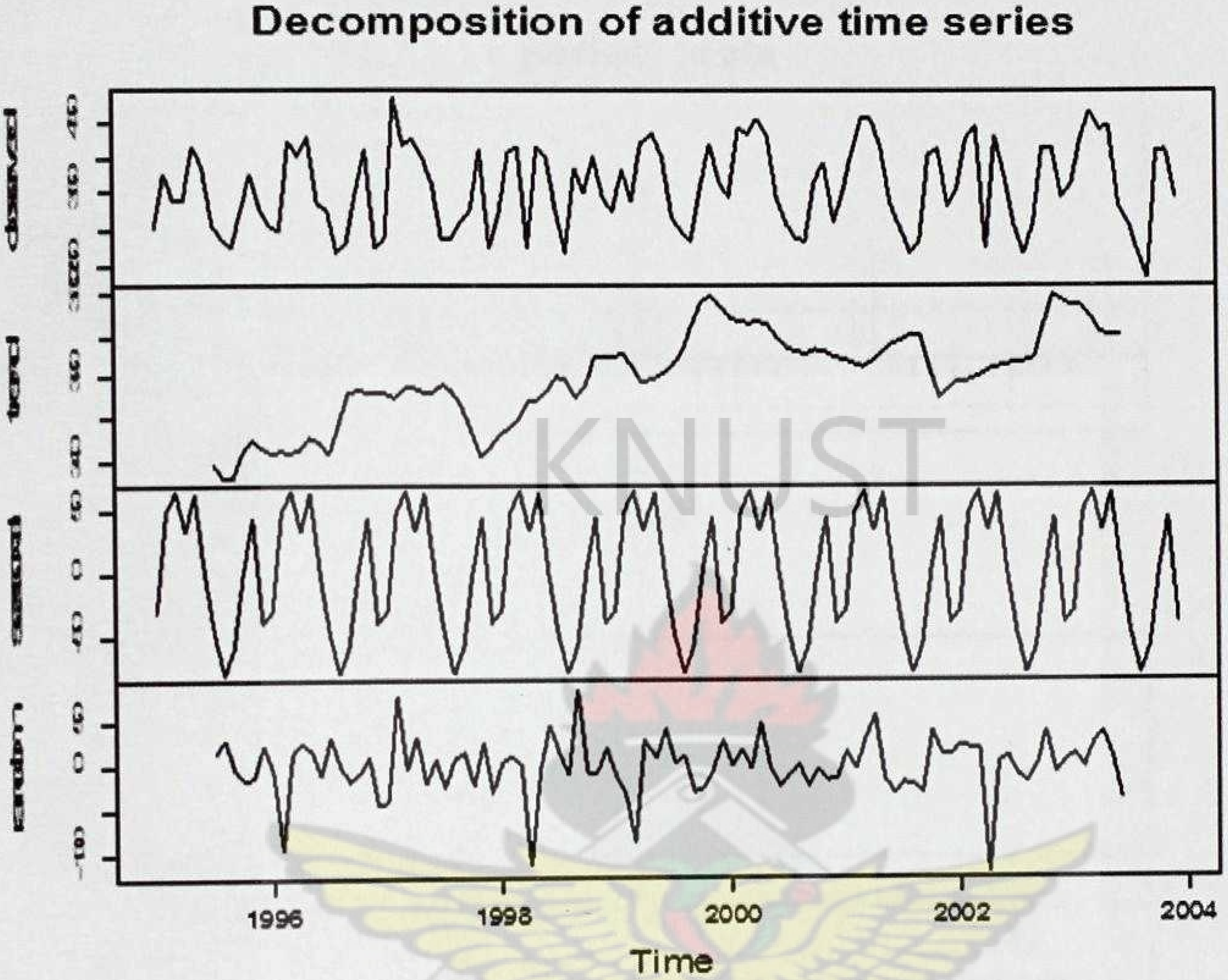

Spectral Density of radiation by Frequency
4.0E5
3.0E5
2.0E5
1.0E5
Density
0.0
0.1
0.2
0.3
0.4
0.5
Frequency
Window: Tukey-Hamming (5)

APPENDIX 8

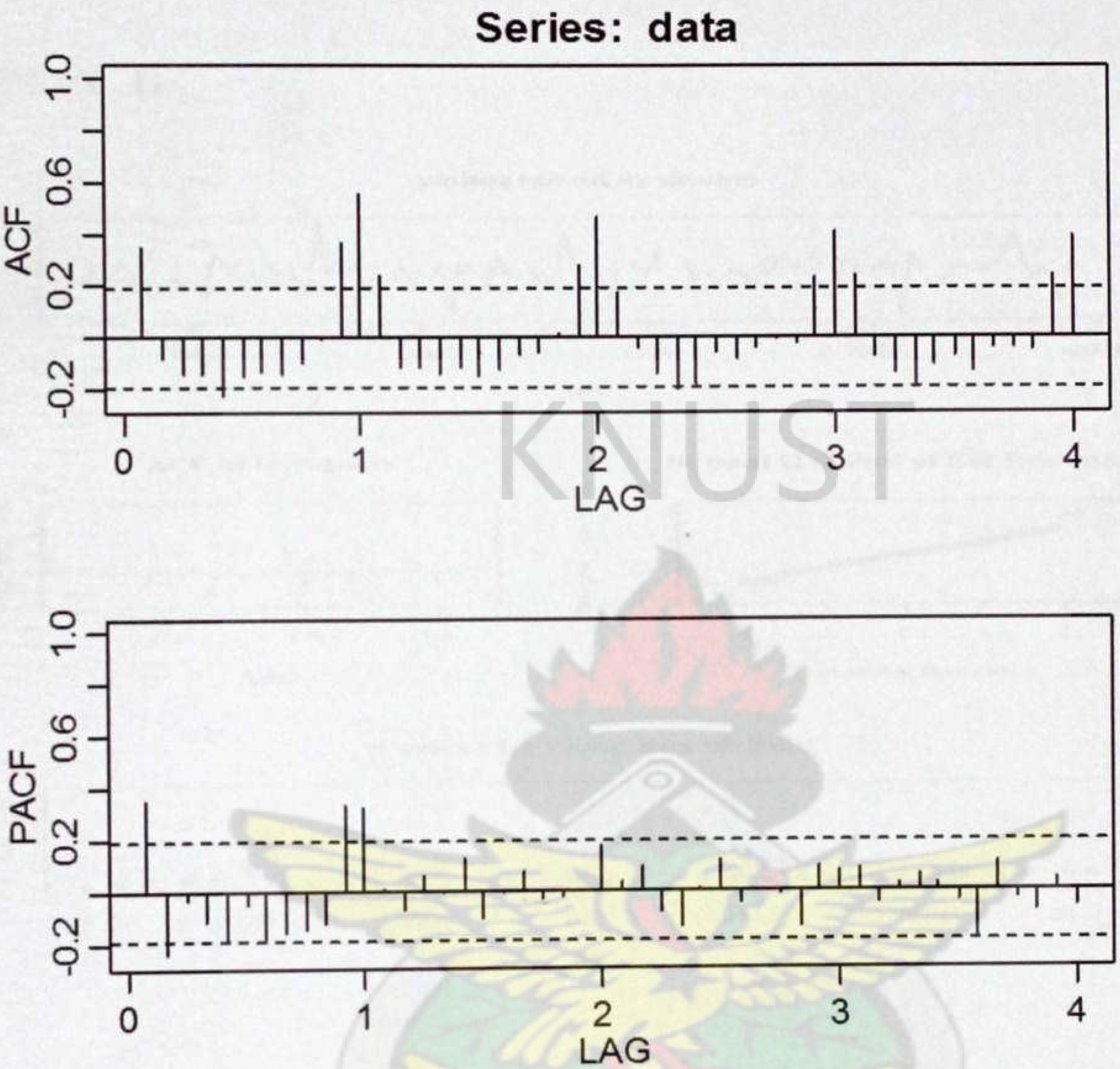


Seasonal adjusted series for radiation from SEASON, MOD_12, ADD EQU 12
450.00000
400.00000
350.00000
300.00000
250.00000
200.00000
JAN 1995
MAY 1995
SEP 1995
JAN 1996
MAY 1996
SEP 1996
JAN 1997
MAY 1997
SEP 1997
JAN 1998
MAY 1998
SEP 1998
JAN 1999
MAY 1999
SEP 1999
JAN 2000
MAY 2000
SEP 2000
JAN 2001
MAY 2001
SEP 2001
JAN 2002
MAY 2002
SEP 2002
JAN 2003
MAY 2003
SEP 2003
Date

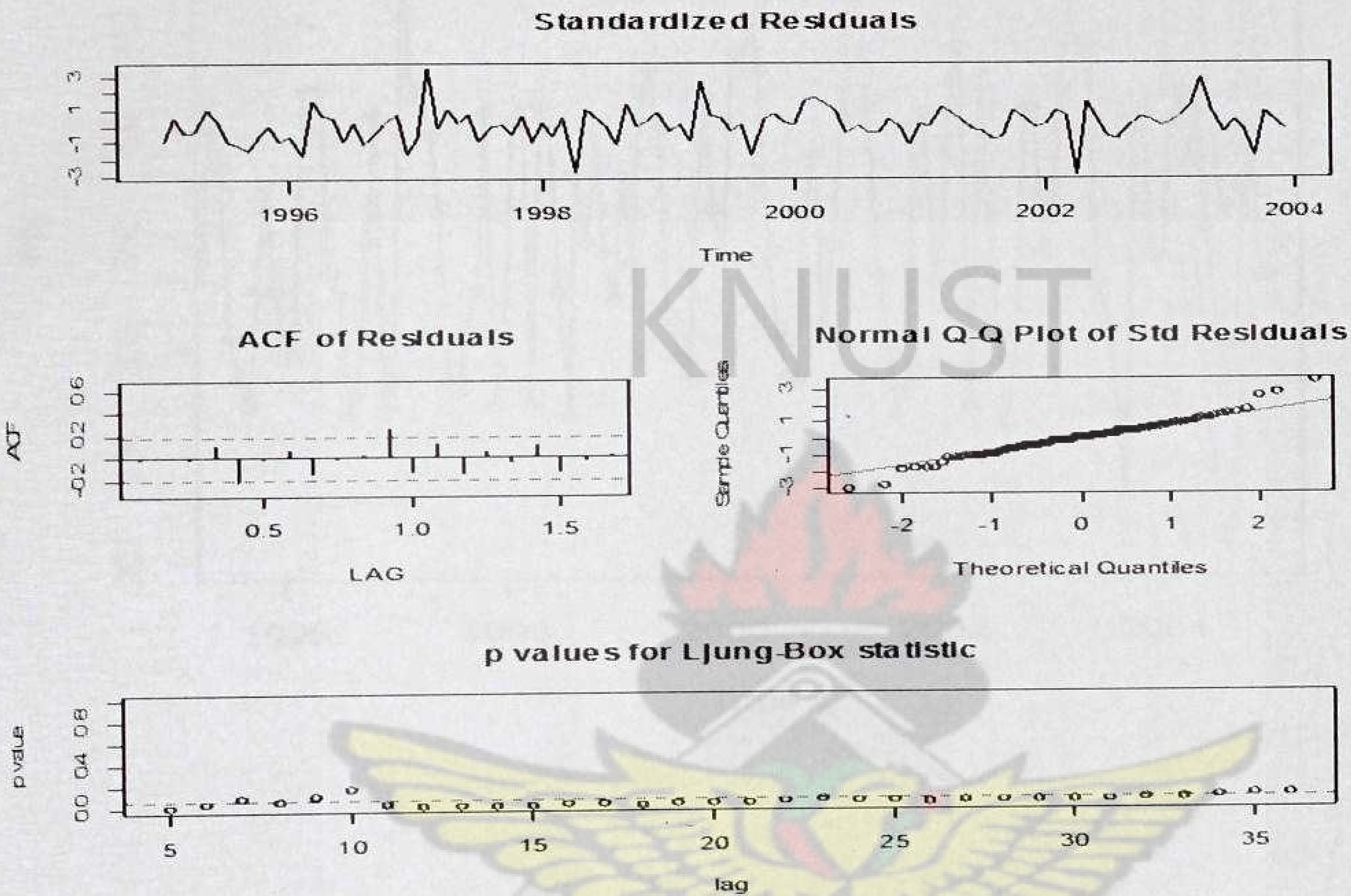
APPENDIX 9


Decomposition of additive time series
1996
1998
2000
2002
2004
Time

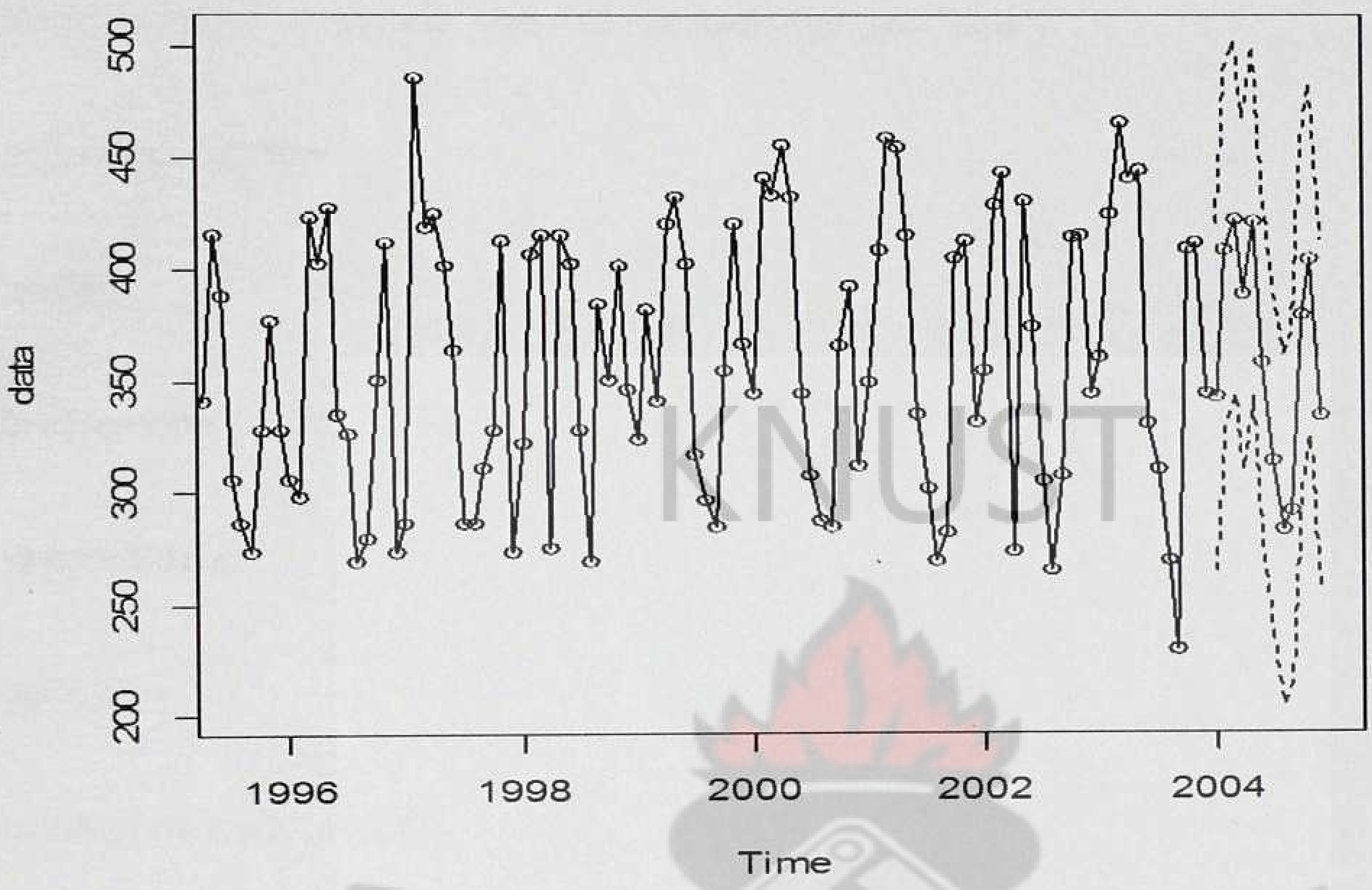
APPENDIX 10



APPENDIX 11


Standardized Residuals
-3 -1 1 3
1996
1998
2000
2002
2004
Time
ACF of Residuals
ACF
-0.2 0.2 0.6
0.5
1.0
1.5
LAG
Normal Q-Q Plot of Std Residuals
Sample Quantiles
-3 -1 1 3
-2
-1
0
1
2
Theoretical Quantiles
p values for Ljung-Box statistic
p value
0.0 0.4 0.8
5
10
15
20
25
30
35
lag

APPENDIX 12



R -CODES

```
data=ts(radiation, start=1995,freq=12)

plot(data,xlab="Time(years)",ylab="Global Radiation",col="blue")

a=decompose(data)

plot(a)

b=a$seasonal

kpss.test(data)

acf2(data)

sarima(data,2,0,1,1,0, 1,12)

sarima(data,0,0,1,1,0,1,12)

sarima(data,2,0,1,1,0,0,12)

sarima.for(data,12,0,0,1,1,0,1,12)
```