INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

Time Series Analysis and Forecasting of Monthly Maximum Temperatures in South Eastern Nigeria

Chisimkwuo John
Department of Statistics, Michael Okpara University of Agriculture, Umudike, Nigeria
Uchechukwu George
Department of Statistics, Michael Okpara University of Agriculture, Umudike, Nigeria
Okezie Sampson Chukwuemeka
Meteorological Department, National Root Crop Research Institute (NRCRI), Umudike, Nigeria

Abstract:
This study examines 1977 – 2012 monthly temperature data for South Eastern Nigeria collected using metrological instrument in the NRCRI Umudike at latitude 050, 291N and longitude 070, 331E (122M above sea level). A preliminary check on the time series plot of the data showed seasonal variation suggesting that the series was not stationary. The classical Box and Jenking's Time Series methodology with its indicative ACF and PACF identification guide was employed. The SARIMA (0, 0, 2) (2, 1, 1)12 model was found to be adequate for the series and monthly forecast from 2013 to 2017 showcased relatively stable temperature values within these years. Verification of the model using the 2011 – 2012 monthly data shows that the model is parsimoniously equitable. At the end, it was recommended that studying and carefully applying mathematical models could help track future rise in monthly temperatures.

Key words: *Temperature, Time Series, Seasonal ARIMA, ACF, PACF, Forecasting*

1. Introduction

The earth's climate system and its changes control life on earth and have a substantial influence on the society and economy. Climate change can be due to natural and anthropogenic effects. It is very important that these can be separated (Volker and Ingeborg, 2006). Today, the entire global community has started suffering from the unfriendly climatic condition, the gradual disappearance of rain forest in the tropics, the loss of plant and animal species, changes in rainfall patterns, and global warming resulting from climate change. Climate change has the potential to affect all natural systems, thereby becoming a threat to human development and survival socially, politically, and economically (Oluwafemi *et al*, 2011).

Climate change is in many countries in the world, one of the biggest environmental threats to food production, water availability, forest biodiversity and livelihoods (Nury *et al*, 2012). Moreover, it is widely believed that developing countries in tropical regions of the world, e.g. Nigeria, will be impacted more severely than developed ones. The climate of Nigeria is tropical however; there are wide climatic variations in different regions of the country. Near coast temperatures rarely exceeds $32^0C(90^0F)$ but humidity is very high and nights are very hot. In land there are two different seasons. A wet season from April to October with lower monthly temperatures and a dry season from November to March with a midday temperatures that rises above $38^0C(100^0F)$ but relatively cool .hts, dropping as low as $12^0C(54^0F)$.

Understanding the nature and scale of possible climate changes in South Eastern Nigeria using the temperature data from meteorological unit of National Root Crop Research Institute, (NRCRI) Umudike is of importance to agricultural production and human health. For this purpose time series analysis of weather data can be a valuable tool to investigate its variability pattern and, maybe, even to predict short and long term changes in the time series.

Time series analysis and forecasting have become a major tool in numerous hydro-meteorological applications to study trends and variations in variables like rainfall, humidity, temperature stream flow and many other environmental parameters.

It is therefore important to forecast how quickly the temperatures are going to increase as temperatures in excess of 40^0C may predispose both human and livestock to heat related diseases (Agrometeorological Bulletin, 2010).

2. Methodology

2.1. The Temperature Data Set

From 1977 to 2012, monthly minimum and maximum temperature data covering the Umudike district has been collected from the NRCRI Umudike Meteorological Department. Temperature data within the district are gathered with the help of meteorological instruments in the NRCRI Umudike at latitude 05^0, 29^1N and longitude 07^0, 33^1E (122M above sea level).

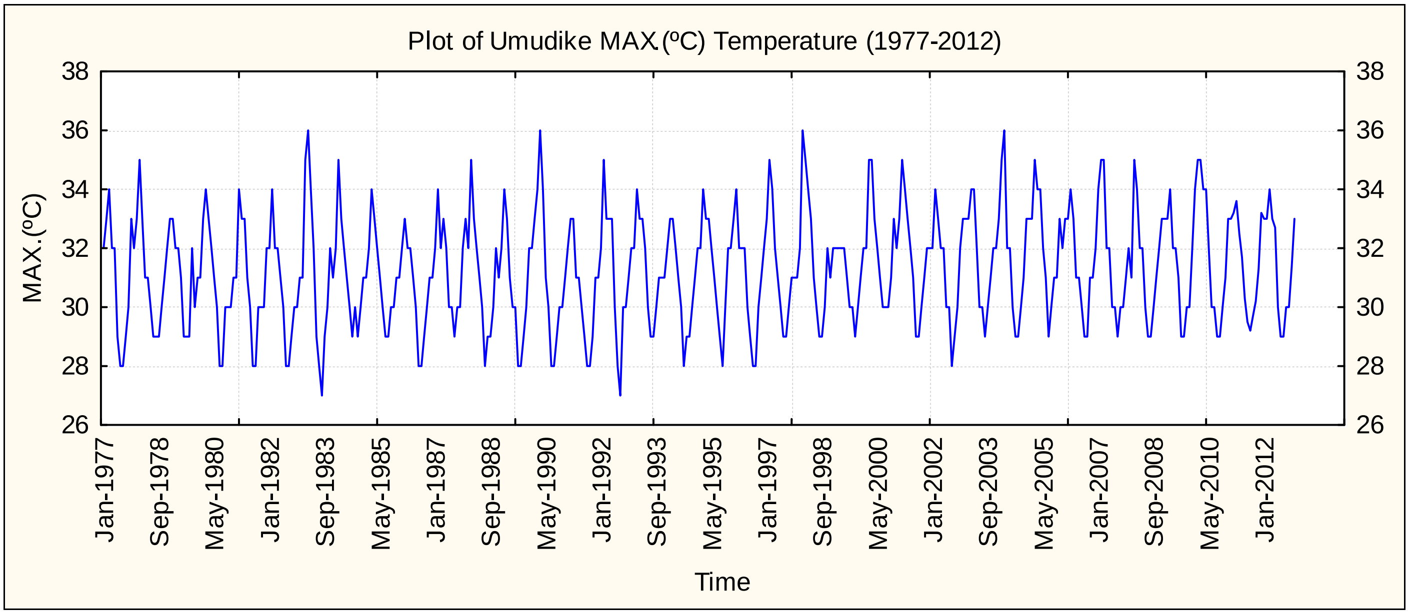

Figure 1: Time Series Plot of Maximum Monthly Temperatures (NRCRI 1977-2012).

An offline descriptive statistics of the data containing the Minimum Eastern region monthly temperature shows that the minimum temperature within the region over the time investigated has not been lower than $17^0\mathrm{C}$. This might not be of great concern to habitants within the Eastern environment and thus, will not form the focus of this study. However, the maximum monthly temperatures in excess of the room temperature might pose a possible climatic hazard to the environment. Consequently, the maximum monthly temperature would be considered in this study since it represents the upper extreme climatic factor. The study attempts to analyse the trend in the maximum monthly temperatures and the forecast as showcased in Figure 1 after identifying a suitable model. A non stationarity situation is observable here as the plot shows a trace of seasonal variation.

2.2. The ARIMA Model

ARIMA is an acronym that stands for Auto-Regressive Integrated Moving Average. This is a known time series model and could be defined algebraically by

$$y_t = \mu + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{p-1} + e_t - \delta_1 e_{t-1} + \ldots - \delta_q e_{t-q} \qquad (1)$$

at time $t = 1, \ldots, n$, where $e_{t-j}\left(j = 0,1,\ldots,q\right)$ are the lagged forecast errors. Usually, the $p+q+1$ unknown parameters $\mu,\ \alpha_1,\ \ldots\alpha_p$ and $\delta_1,\ \ldots\delta_q$ are determined by minimizing the squared residuals (Box and Jenkins, 1976).

From the ARIMA technique, the dependent variable y_t is predicted in the first part of the right hand side of equation (1) above based on its values at earlier time periods. This constitutes the autoregressive (AR) part in equation (1) above. In the second part, the dependent variable y_t also depends on the values of the residuals at earlier time periods, which may be regarded as prior random alarms. This is the moving average (MA) part of equation (1).

In addition to the AR and MA parameters, ARIMA models may also include a constant. The interpretation of a (statistically significant) constant depends on the model that is fit. Two indicative situations are:

- The situation of no autoregressive parameters in the series. In such case, the expected value of the constant is μ, the mean of the series;
- The situation of autoregressive parameters in the series. In such case, the constant represents the intercept. If the series is differenced, then the constant represents the mean or intercept of the differenced series. For the non-seasonal scenario, the simple $\mathrm{ARIMA}(p,d,q)$ model is used with p the number of autoregressive terms, d the number of non-seasonal

differences, and q the number of lagged forecast errors in the prediction equation. However, climatic data usually contains the seasonal variations. Thus, it is more apt to incorporate the full Seasonal Auto Regressive Integrated Moving Average (SARIMA) model

$$\text{SARIMA}(p,d,q)(P,D,Q)_S \qquad (2)$$

With P the order of the seasonal AR-model; D the order of the seasonal differencing and Q the order of the seasonal MA-model. The subscript s is the number of periods in the season. Mathematically, the general form of the model represented in equation (2) above can be written in the backshift notation (B) as

$$\alpha_{AR}(B)\alpha_{SAR}\left(B^S\right)(1-B)^d\left(1-B^S\right)^D y_t = \delta_{MA}(B)\delta_{SMA}\left(B^S\right)e_t \cdots \qquad (3)$$

where α_{AR} is the non-seasonal AR parameter, δ_{MA} the non-seasonal MA parameter, α_{SAR}, the seasonal AR parameter, and δ_{SMA} the seasonal MA parameter.

2.3. The Stationarity Condition

Stationarity is an important condition for ARIMA models. In practice, the mean and variance should be constant as a function of time before performing the analysis. Otherwise, past effects would accumulate and the values of successive y_t 's would approach infinity making the process non-stationary. For a first order non-stationarity, the observations with ARIMA models should be sieved first by differencing the observations d times, using $\Delta^d y_t$ instead of y_t as the time series to obtain stationary data. This is usually done with the transformation

$$\Delta y_t = y_t - y_{t-1} \qquad (4)$$

The operations of equation (4) will result to the values $d = 0,1,2,...$ for the non-seasonal part and values $D = 0,1,2,...$ for the seasonal part and this serves as an indicative guide in eliminating the first order non-stationarity in the model identification process. Note that in the situation of a second order non-stationarity, a simple transformation (e.g. the log transformation) could be a worthwhile procedure to apply when detected.

3. Applying the Arima Technique

So far, the above centered on the Box and Jenkins methodology with its benchmark model application procedure on mainly three steps (a) identification (b) estimation (c) and the forecasting or diagnostic checking. The identification stage involves the determination of tentative values of the p,d,q and the P,D,Q sets using the linear least squares method. In the identification stage, a stationary or a weakly stationary situation is obtained by differencing and transformation of the data if needed. Then, the ACF and the PACF plots are used to suggest possible models by determining the orders p and q in the Seasonal $\text{ARIMA}(p,d,q)(P,D,Q)_S$ model. The goodness of the best models could be evaluated using the Mean Square Error (Residuals) MSE or using the Akaike Information Criterion AIC (Tanja, 2010).

3.1. Preliminary Check and Model Detection

A SAR (1) estimate of 0.99987, which is closer to one strongly suggests that the model is non-stationary at this stage and might need to be differenced at the seasons. The left and right panels of Figure 2 shows the ACF and PACF of the differenced SAR (1) preliminary analysis of the maximum temperatures respectively up to lag 36. Observe that the ACF (Left Panel) seems to cut off after the first season suggesting a model with $Q = 1$. The PACF seems not to have significant lags extending the first two or three seasons suggesting an initial model gaze with $P = 2$.

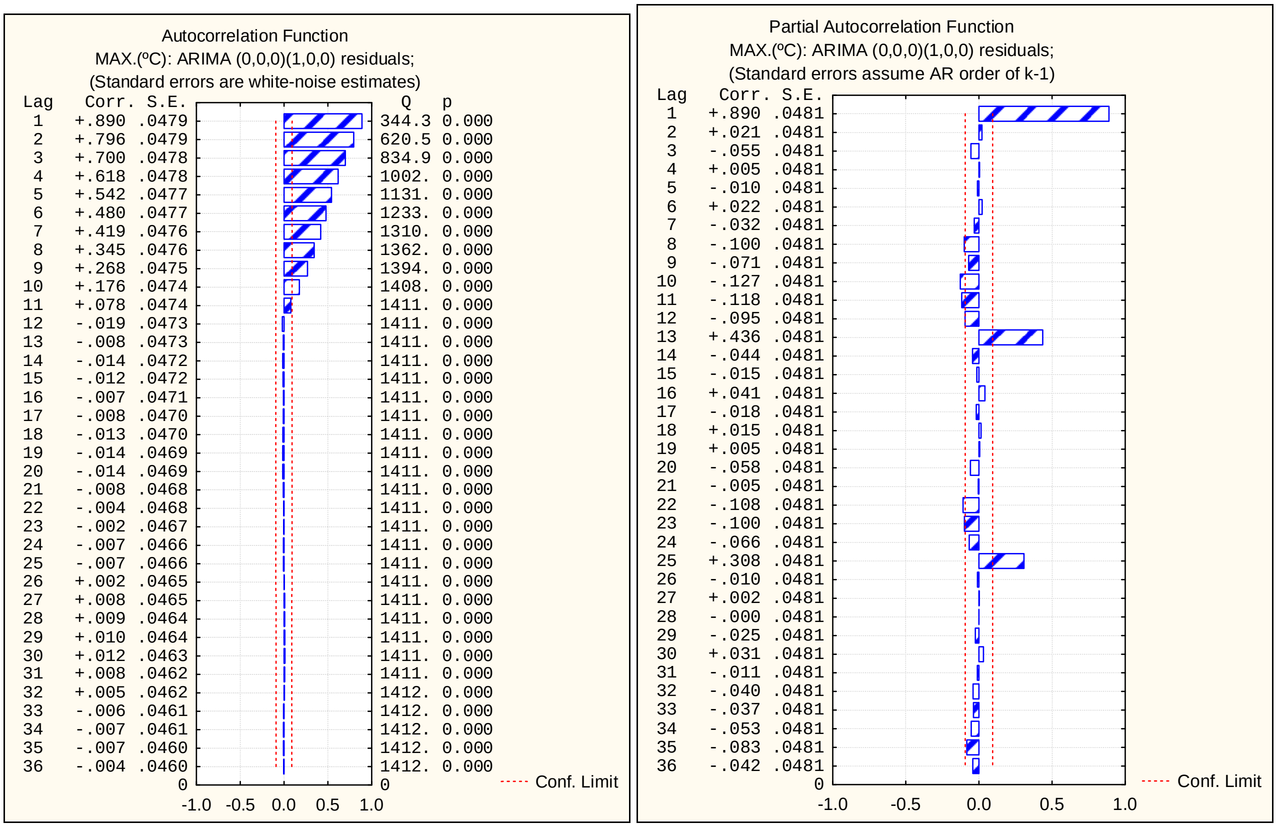

Figure 2: ACF (Left Panel) and PACF (Right Panel) of SAR (1) Residuals – NRCRI Temp. Data

So far, the study has detected a seasonal effect s, an indicative $Q=1$, $P=2$ from the ACF and PACF respectively, and a parsimonious $D=1$ from the first seasonal difference. Thus, tentative models obtainable from the ongoing preliminary analysis are shown in Table 1.

MODEL	MSE	
	Conditional Method	Unconditional Method
SARIMA(0,1,1)(1,0,0)$_{12}$	1.12270	1.10060
SARIMA(1,1,0)(1,1,0)$_{12}$	1.03510	1.02660
SARIMA(2,1,0)(1,0,0)$_{12}$	1.13760	1.11850
SARIMA(1,0,2)(1,1,0)$_{12}$	0.78134	0.77491
SARIMA(0,0,1)(1,1,0)$_{12}$	0.78900	0.78302
SARIMA(0,0,1)(2,1,1)$_{12}$	0.62063	0.56184
SARIMA(0,0,2)(2,1,1)$_{12}$	0.61157	0.55127
SARIMA(0,1,2)(1,1,0)$_{12}$	0.84553	0.79193

Table 1: SARIMA Models for the NRCRI Maximum Temperatures (1977-2011)

Based on the MSE of the various indicative Seasonal ARIMA models displayed in Table 1, the SARIMA(0,0,2)(2,1,1)$_{12}$ using the unconditional method (STATISTICA Help Documentations, 2010) could be chosen as the best fitted model for the NRCRI (1977-2012) maximum temperature dataset.

3.2. Model Adequacy

The identified ARIMA model must be diagnostically checked for its appropriateness or adequacy. The Box and Jenkings prescribed approach to this entails looking at the ACF and PACF of the model's residuals. These are shown for the monthly maximum temperatures at the NRCRI station in Figure 3. Observe that the spikes at the ACF (left panel) and the PACF (right panel) up to lag 36 are within the 95% statistical confidence bounds respectively. This suggests that the SARIMA$(0,0,2)(2,1,1)_{12}$ is adequate for the NRCRI (1977-2012) maximum temperature time series dataset.

The results obtained from the ACF and PACF of Figure 3 is supported by the residual series plotted in Figure 4 as the values centers randomly around the zero value. This is readily observable in Figure 4 when traced from the Right Y-axis.

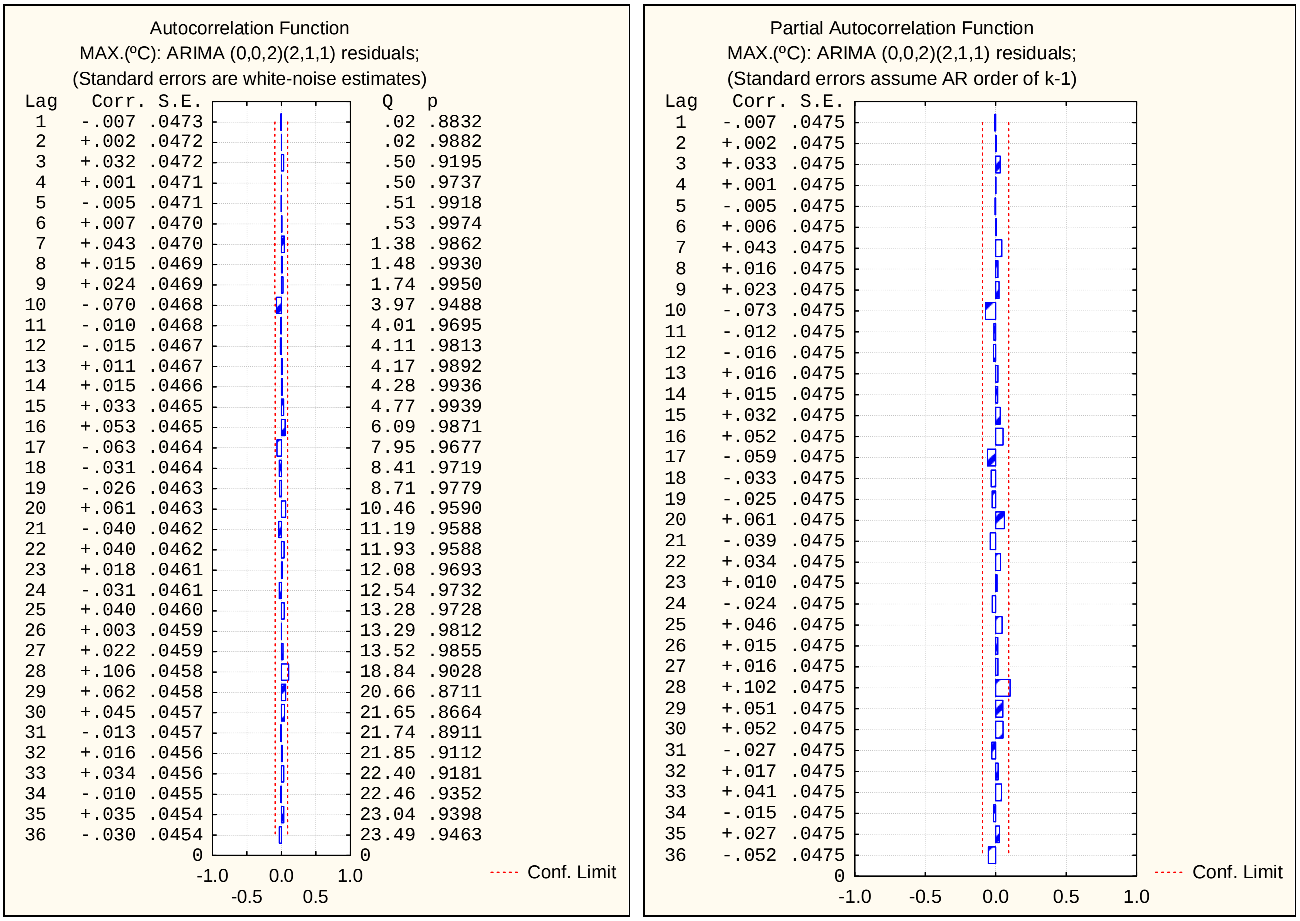

Figure 3: ACF(Left Panel) and PACF(Right Panel) of SARIMA$(0,0,2)(2,1,1)_{12}$ – NRCRI Temp. Data

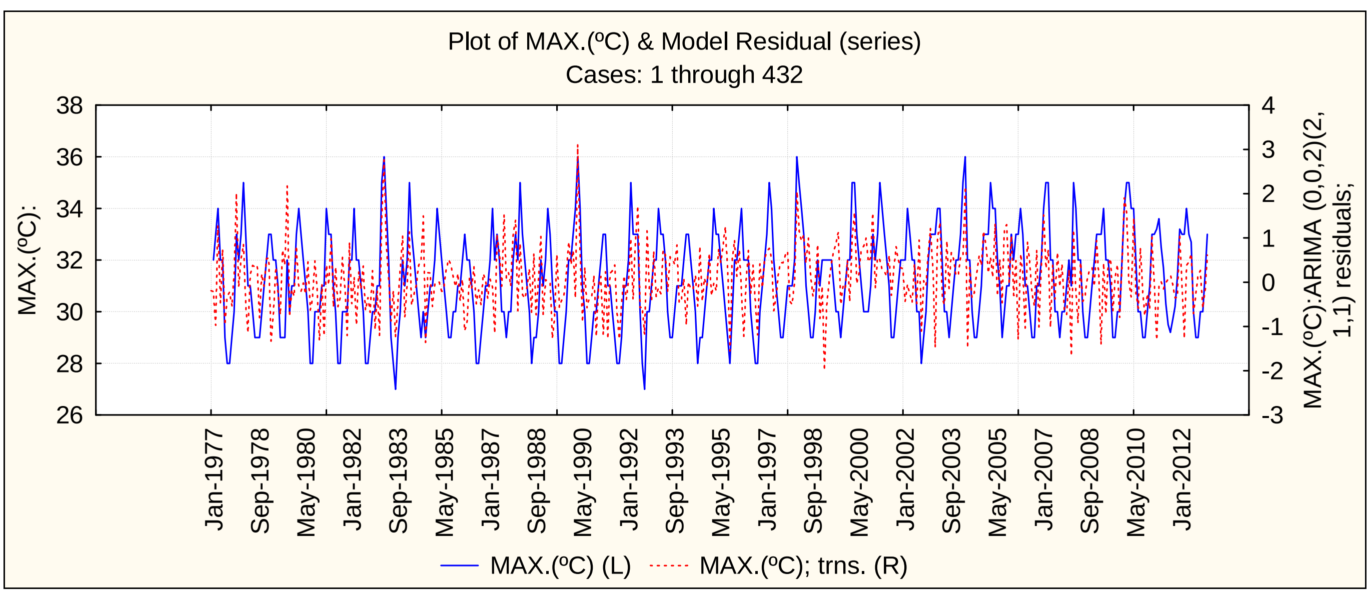

Figure 4: Original Series (Left Y-axis) and SARIMA $(0, 0, 2) (2, 1, 1)_{12}$ Residual Series (Right Y-axs)

3.3. Forecasting

A final and another useful step in the Box and Jenkin's approach is the application of the identified model in forecasting one or more future time steps ahead. Using the identified model's parameters, Figure 5 shows the one-month-ahead predictions with their 95% confidence limits for the monthly maximum temperatures for the next five years (2013 – 2017) at the NRCRI station. The years 2011-2012 was used to verify the predictability of the model as exhibited in Figure 6.

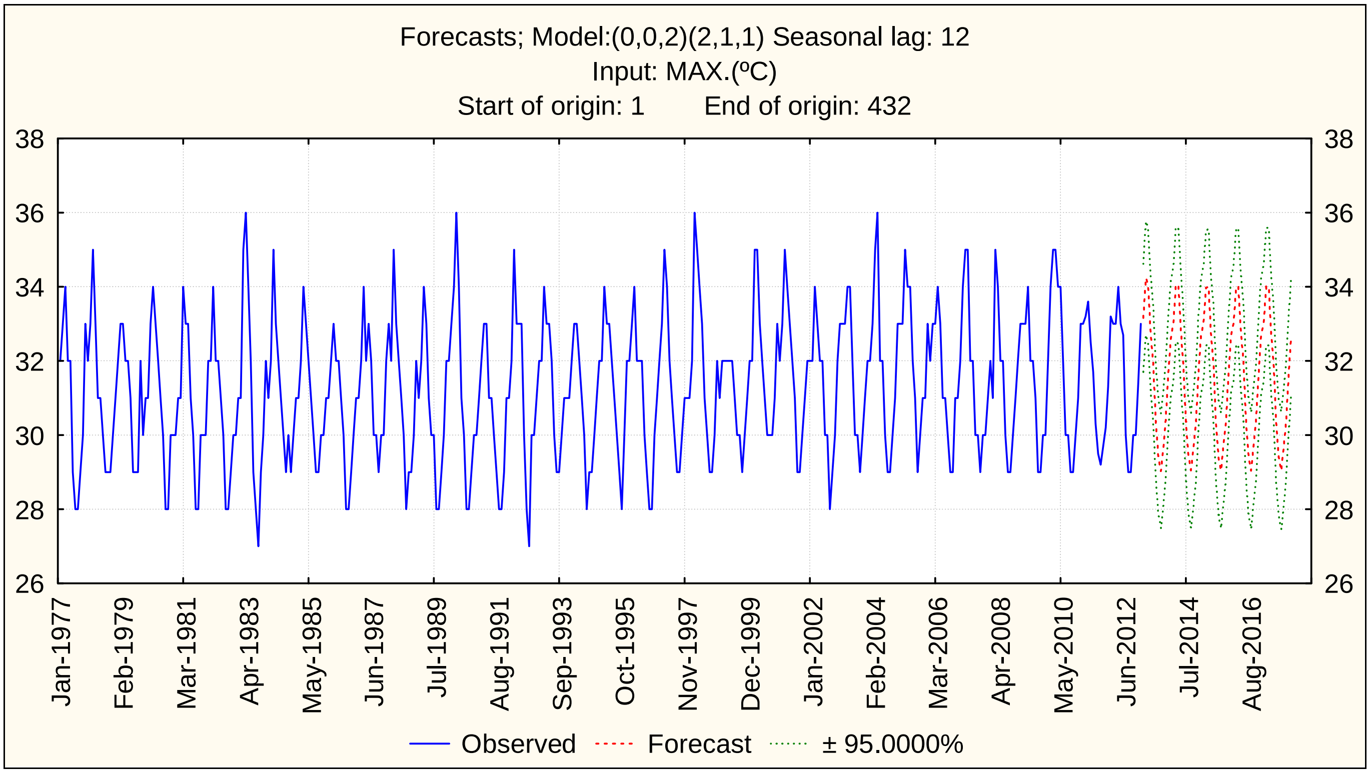

Figure 5: Original Plot with five Years Forecast at 95% Confidence Interval

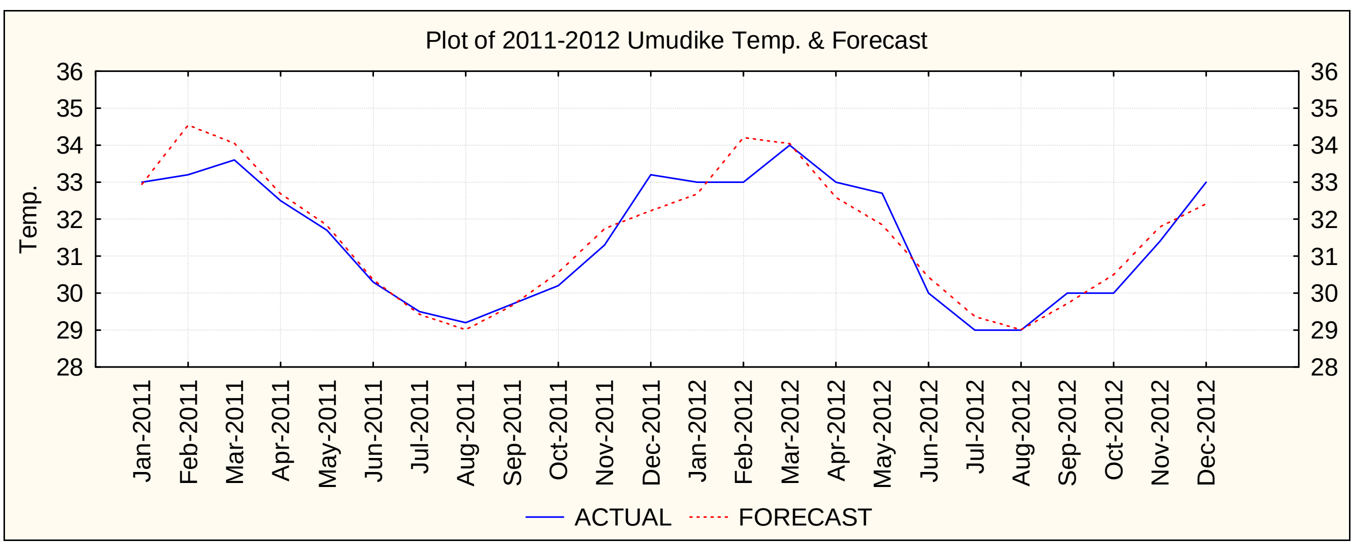

Figure 6: 2011-2012 Original Data and Predicted Data

4. Conclusion

The Box-Jenkins ARIMA methodology has proved to be a useful technique that can help decision makers to establish better strategies and to set up priorities for equipping themselves against upcoming weather changes. The maximum temperature time series fitted by this procedure for the NRCRI station shows that it is possible to predict the evolution of the maximum temperature of Eastern Nigeria based on the data collected for the past 36 years. Based on the best suited model, the maximum temperature for the next five years seems to be slightly stable from that of the reference period 1977-2012, with maximum values not exceeding the period's maximum. Of course, this approach has shown dependable results and a general recommendation is that careful understudy of various mathematical models like the ARIMA could help track future rise in monthly temperature although for a relatively short time intervals. Indeed as we advance in time, the uncertainty about the predictions grows, so the results might become indecisive.

5. References

1. BOX, G.E.P. & JENKINS, G.M. (1976), Time Series Analysis: Forecasting and Control. Revised Edition, Holden-Day: San Francisco, CA.
2. NIGERIA METEOROLOGICAL AGENCY (2010), Agrometeorological Bulletin, NIMET Meteorological Agency Abuja-Nigeria.
3. NURY, A.H., KOCH, M. & ALAM, M. J. B. (2013), Time Series Analysis and Forecasting of Temperatures in Sylhet Division of Bangladesh, 4th International Conference on Environmental Aspects of Bangladesh (ICEAB), August 24-26.
4. OLUWAFEMI, S.O., FEMI, J.A. & OLUWATOSI, T.D. (2010), Time Series Analysis of Rainfall and Temperature in South West Nigeria, the Pacific Journal of Science and Technology, Vol.11 No.2 pp 552-564.
5. STATSOFT, INC (2010), STATISTICA Version 9.1 Help Documentations, URL http://www.statsoft.com/.
6. TANJA VAN HECKE, (2010), Time Series Analysis to Forecast Temperature Change, Ghent University Academic Bibliography 35(2). P.63-69.
7. VOLKER, W. & INGEBORG, H. (2006), The Climate Station of the University of Hohenheim: Analyses of Air Temperature and Precipitation Time Series since 1878; International Journal of Climatology, Vol.26, pp 113-138