MODELLING OF DAILY TEMPERATURE OF SOKOTO TOWN USING BOX-JENKINS PROCEDURE

Yakubu M. Yeldu[1], Shehu L.[2], Mukhtar G[3] And M.b. Shehu[4]
[1,2,3,4]Department of Statistics, College of Science and Technology
Waziri Umaru Federal Polytechnic, Birnin Kebbi

Abstract

This paper aims to model the daily temperature of Sokoto town in North western Nigeria using the Box-Jenkins methodology. Daily temperature for the period 1st January 2009- 3rd August, 2013 was used. The empirical study reveals that the most adequate model for the daily temperature of Sokoto town for the period under study is ARIMA (3, 1, 1). Using the model, we forecast the values of daily temperature for the next fourteen days i.e. 4th -17th of August, 2013. Results obtained revealed decreasing trend for the first ten days but started to increase towards the end.

Keywords: *Daily temperature, Box-Jenkins methodology; ARIMA models, Forecasting, Sokoto*

Introduction

Sokoto, the capital of Sokoto state in North West of Nigeria is in the dry Sahel surrounded by sandy savannah and isolated hills. With an annual average temperature of 28.3 °C (82.9 °F), Sokoto is one of the hottest cities in the world, however the maximum daytime temperatures are most of the year generally under 40 °C (104.0 °F), and the dryness makes the heat bearable (Tsoho, 2008). The warmest months are February to April, where daytime temperatures can exceed 45 °C (113.0 °F). Highest recorded temperature is 47.2 °C (117.0 °F), which is also the highest recorded temperature, in Nigeria (Wikipedia, 2013). The rainy season is from June to October, during which showers are a daily occurrence. The showers rarely last long and are a far cry from the regular torrential showers known in many tropical regions. From late October to February, during the 'cold season', the climate is dominated by the Harmattan wind blowing Sahara dust over the land. The dust dims the sunlight, thereby lowering temperatures significantly and also leading to the inconvenience of dust everywhere in the house. The region's lifeline for growing crops is the floodplains of the Sokoto-Rima river system, which are covered with rich alluvial soil. For the rest, the general dryness of the region allows for few crops, millet perhaps being the most abundant, complemented by maize, rice, other cereals, and beans (Sokoto Diary, 2012). Apart from tomatoes, few vegetables grow in the region. The low variety of foodstuffs available has resulted in the relatively dull local cuisine. In terms of vegetation, Sokoto falls within the savannah zone. This is an open Tse-tse fly free grassland suitable for cultivation of grain crops and animal husbandry. Rainfall starts late and ends early with mean annual falls ranging between 500 mm to 1,300 mm. There are two major seasons in Sokoto namely wet and dry. The dry season starts from October, and lasts up to April in some parts and May extend to May or June in other Parts. The wet season on the other hand begins in most parts of the state in May and lasts up to September, or October. The harmattan, a dry, cold and fairly dusty wind is experienced in the state between November and February. Heat is more severe in the state in March and April. But the weather in the state is always cold in the morning and hot in the afternoons save in peak at harmattan period. The topography of the state is dominated by famous Hausa plain of northern Nigeria. The vast fadama land of the Sokoto-Rima River systems dissects the plain and provides the rich alluvial soil fit for variety of crops cultivation in the state. There are also isolated hills and mountains ranges scattered all over the state.

Background of the Study

When sunlight reaches Earth's surface some is absorbed and warms the earth and most of the rest is radiated back to the atmosphere at a longer wavelength than the sun light. Some of these longer wavelengths are absorbed by greenhouse gases in the atmosphere before they are lost to space. The absorption of this longwave radiant energy warms the atmosphere. These greenhouse gases act like a mirror and reflect back to the Earth some of the heat energy which would otherwise be lost to space. The reflecting back of heat energy by the atmosphere is called the "greenhouse effect". Global warming is the observed and projected increases in the average temperature of Earth's atmosphere and oceans. The higher the concentration of green house gases like carbon dioxide in the atmosphere, the more heat energy is being reflected back to the Earth. (Global warming art, 2013) The emission of carbon dioxide into the environment mainly from burning of fossil fuels (oil, gas, petrol, kerosene, etc.) has been increased dramatically over the past 50 years (Time for Change, 2013).

Increasing global temperatures are causing a broad range of changes. Sea levels are rising due to thermal expansion of the ocean, in addition to melting of land ice. Amounts and patterns of precipitation are changing. The total annual power of hurricanes has already increased markedly since 1975 because their average intensity and average duration have increased (in addition, there has been a high correlation of hurricane power with tropical sea-surface temperature). Changes in temperature and precipitation patterns increase the frequency, duration, and intensity of other extreme weather events, such as floods, droughts, heat waves, and tornadoes. Other effects of global warming include higher or lower agricultural yields, further glacial retreat, reduced summer stream flows, species extinctions. As a further effect of global warming, diseases like malaria are returning into areas where they have been extinguished earlier.

Almost 100% of the observed temperature increase over the last 50 years has been due to the increase in the atmosphere of greenhouse gas concentrations like water vapour, carbon dioxide (CO_2), methane and ozone (Time for change, 2013). Although most studies focus on the period up to 2100, warming is expected to continue past then because CO_2 has an estimated atmospheric lifetime of 50 to 200 years. According to different assumption about the future behaviour of mankind, a projection of current trends as represented by a number of different scenarios gives temperature increases of about 3° to 5° C (5° to 9° Fahrenheit) by the year 2100 or soon afterwards. A 3°C or 5° Fahrenheit rise would likely raise sea levels by about 25 meters (about 82 feet).Global warming art,2013)

Objectives of the Study

The aim of this paper is to study the pattern and behaviour of the daily temperature of Sokoto town in north western Nigeria which is among the highest in the country. To achieve this, the following objectives are formulated:

I. To develop Autoregressive integrated moving average (ARIMA) models for the daily temperature of Sokoto town.
ii. To identify among the models the one that best fit the data under study.
iii. Use the model to forecast future temperature of Sokoto town.

Data and Methodology

The data used for this study is secondary data on the daily temperature of Sokoto town for the period 1st January, 2009 to 3rd of August, 2013. It was collected from the records of energy commission of Nigeria at the Sokoto energy research centre in Sokoto town.

The methodology adopted is the concept of ARIMA modelling. The pioneers in this area were Box and Jenkins who popularized an approach that combines the moving average and the autoregressive models. Although both autoregressive and moving average approaches were already known (and were originally investigated by Yule), the contribution of Box and Jenkins was in developing a systematic methodology for identifying and estimating models that could incorporate both approaches. This makes Box-Jenkins models a powerful class of models. The Box-Jenkins ARMA model is a combination of the AR and MA models as follows:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} - b_1 u_{t-1} - b_2 u_{t-2} - \dots - b_q u_{t-q} + u_t$$

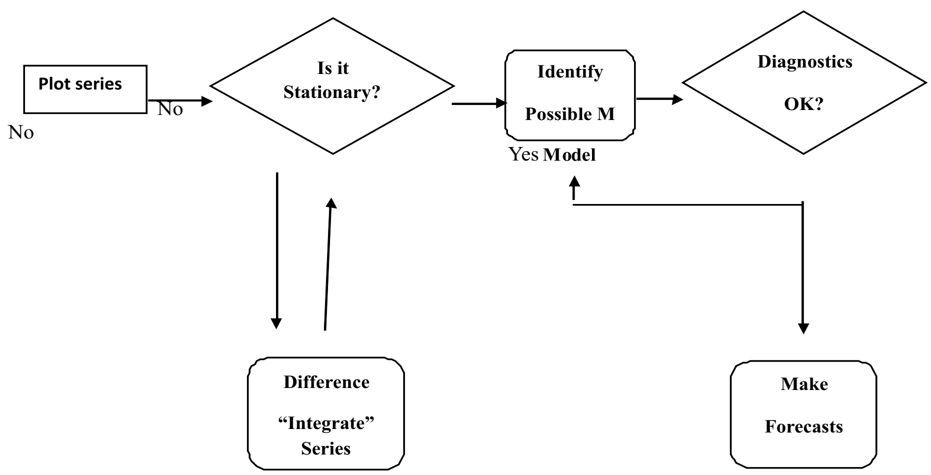

Figure 1. Box-Jenkins procedure

In the Box-Jenkins procedure, there are three primary stages in building a Box-Jenkins time series model:

1. Model Identification
2. Model Estimation
3. Model Validation

The identification stage is the most important and also the most difficult: it consists of determining the adequate model from ARIMA family models. This phase is founded on the study of autocorrelation and partial autocorrelation. The first step in developing a Box-Jenkins model is to determine if the series is stationary and if there is any significant seasonality that needs to be modeled.

The Box-Jenkins model assumes that the time series is stationary. A stationary series has:

1. Constant mean
2. Constant variance
3. Constant autocorrelation structure

Regression with non stationary variables is a spurious correlation. The random walk:

$$y_t = y_{t-1} + u_t,\ u_t \sim N(0, \sigma^2)$$

is not stationary, since its variance increases linearly with time. Stationarity can be accessed from a run sequence plot. The run sequence plot should show constant location and scale. It can also be detected from an autocorrelation plot. Specifically, non-stationarity is often indicated by an autocorrelation plot with very slow decay. Box and Jenkins recommend differencing non-stationary series one or more times to achieve

stationarity. Doing so produces an ARIMA model, with the "I" standing for "Integrated". But its first difference $\Delta y_t = y_t - y_{t-1} = u_t$ is stationary, so y is integrated of order 1", or $y \sim I(1)$.
Many test statistics have been developed to check whether the series contains unit roots or not. The most popular of them is Dickey-Fuller test. Dickey and Fuller (1979) introduced Dickey – Fuller (DF) test statistic to test whether the series contains unit root or not. However, it can be any other process also. Because of this Augmented Dickey–Fuller (ADF) test statistic has developed in the same manner to check the stationarity of the series.

The null hypothesis that the data generating process (DGP) is stationary is tested against a unit root. Kwiatkowski, Philips, Schmidt, and Shin (KPSS) (1992) have yt = Xt + Zt derived a test for this pair of hypothesis. If there is no linear trend term, they start from DGP.
Where, Xt is a random walk i.e. $x_t = x_{t-1} + v_t, v_t \sim iid(0, \sigma_v^2)$ and Z_t is stationary process. Two statistical software packages namely Gretel and J-multi were used in the analysis of the collected data as well as running all the tests on the data.

Analysis and Results
The collected temperature data for the period (1/1/2009-03/08/2013) consists of 1676 daily observations. These were used to build a suitable ARIMA (p, d, q) model to forecast the temperature series over the period of two weeks.

Stationarity test results
The process starts with testing the series for stationarity using the plot diagram and also performing Unit-root test (ADF). The graphical representation of the series against time indicates that the underlying series exhibit an increasing trend over time and has a random walk time series with a non-zero mean and a non-constant variance. Hence, the plotted graph provides a clear cut indication that the underlying series is non-stationary in its level. But the first difference of the series was stationary (Fig.2 and 3).

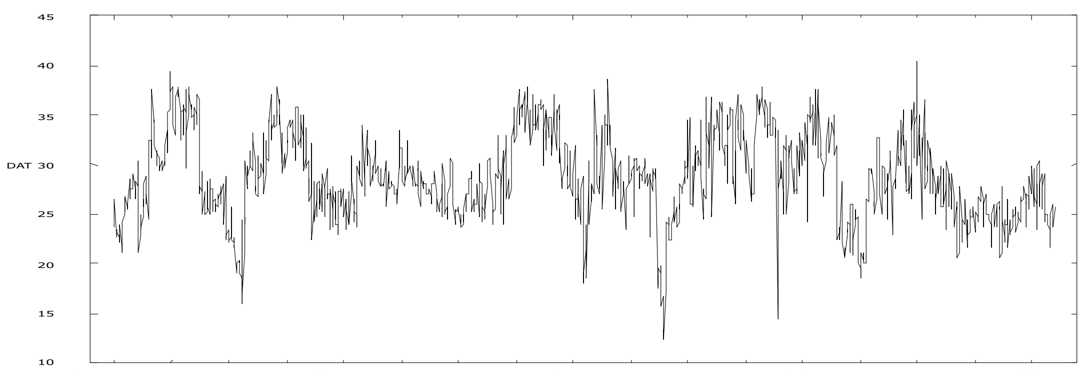

Fig. 2: Graph showing that the data is non stationary

It is revealed from figure 1 above that the daily temperature of Sokoto town for the period under study is non stationary due to an unstable mean which increase and decrease at certain points. The mean and variance thought to be adjusted to form stationary series, so that the values vary more or less uniformly about a fixed level over time. The mean is not constant throughout the series as it assumes a downward trend by decreasing from the highest peak to the lowest peak. There are sudden swings over the days after which the mean stabilizes in the remaining period; hence the mean and variance are non stationary. A normality test performed on the mean and variance using the Anderson-Darling Normality Test at 95% confidence interval figure 3 below revealed that the mean is rightly skewed with a mean value of 28.79699 and variance of 18.003. The coefficient of skewness and kurtosis are -0.060749 and 2.994989 respectively. It is evident at 5% significant level that there are large swings in the data indicating non stationarity.

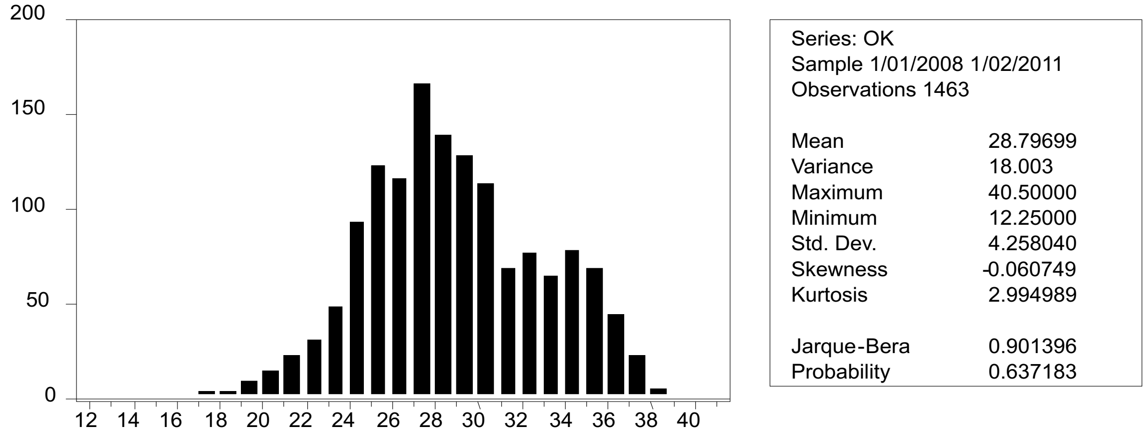

Fig 3: Testing for normality

Due to the non stationary of the data which we observe from the time series plot and the Anderson Darling Normality test we also apply the unit root test which is in table 1. Evidently from figure 4, the variance and mean looks stationary as compared to the figure 2, but this is not enough to test for stationary with respect to the mean and variance. A confirmatory test on stationary is followed through using the Augmented Dickey Fuller unit root test on the first and KPSS Test as shown in Table 2.

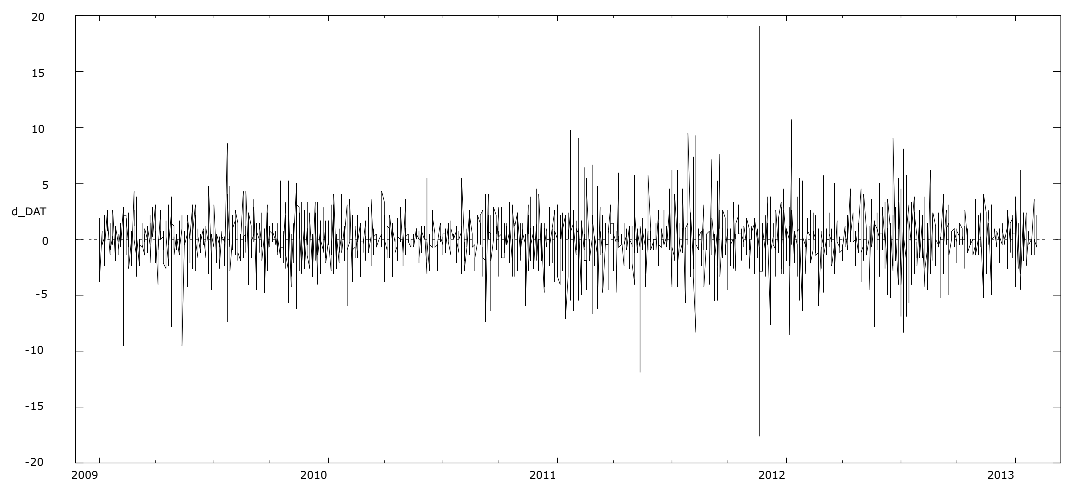

Fig 4: Graph showing the data is stationary after transformation

The advanced analytical technique for testing the stationarity of the time series data uses the Augmented Dickey-Fuller (ADF) test and KPSS test statistics.

The ADF test statistic test the null hypothesis of presence of unit root against the alternative of no unit root and the decision rule is to reject the null hypothesis when the value of test statistic is less than critical value. The KPSS test statistic test the null hypothesis of stationarity against the alternative of unit root and the decision rule is to accept the null hypothesis when the value of the test statistic is less than the critical value. The unit-root test results reject the null hypothesis at 5% level of significance indicating that the series is non-stationary in its level (Table 1). Table 2 presents the unit root test for the first difference of the series. Therefore, one can reject the hypothesis at a 5% percent level of significance that the series is stationary at its first difference form. The temperature series is integrated of order one, (1).

Table 1: Table showing results of further tests

Unit root test	Computed Value	Critical Value at 5%
Augment Dickey – Fuller (ADF)	-1.56478	- 3.41
KPSS Test	0.646226	0.146

Table 2: Test for unit roots for the first difference

Unit root test	Computed Value	Critical Value at 5%
Augment Dickey – Fuller (ADF)	-12.5803	-2.89
KPSS Test	0.1089	0.463

KPSS Test0.10890.463The output of the Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) in figure 5 respectively indicates that both ACF and PACF tail off after cutting at lag two, thus they both tend to decay consistently as the lags die out at the later part of the output. The ACF cuts at lag two (2), there after it decays and dies off at the later Part of the output in figure 3. The PACF also cuts at lag twelve (12), after which it decays and dies off at the later part of the output. This suggests the use of an Autoregressive Integrated moving-average of order (p, d, q). On this account several models such as ARIMA (1, 1,1), ARIMA (2, 1, 1) and ARIMA (1, 1, 0) models are suggested for tentative model selection.

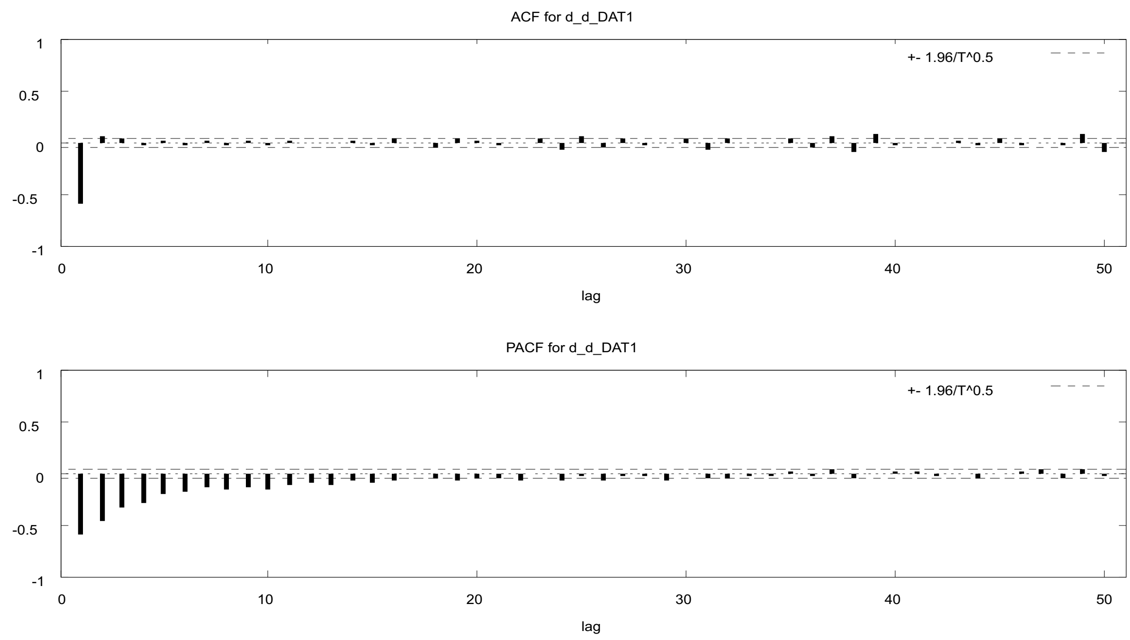

Fig 5: Autocorrelation and partial autocorrelation functions

Model Fitting

After the series become stationary by taking the first difference, the first step is to identify the order of both the AR and MA parts of the ARIMA model. Based on the first difference order to be (1), different forms of ARIMA models were suggested as the following: ARIMA (3.1.2), ARIMA (3.1.1), and ARIMA (2.1.2). The procedure of choosing the most suitable model relies on choosing the model with the minimum AIC and BIC criteria. It can be seen that ARIMA (3, 1, 1) is the best model (Table 3).

Table 3: Comparism of AIC and BIC

MODEL	AIC	BIC
ARIMA (3,1,2)	598.9426	616.3296
ARIMA (3,1,1)	596.5857	611.0749
ARIMA (2,1,2)	605.2186	619.7078

Diagnostic checking

Once the ARIMA model is identified, the test of the suitability of the selected ARIMA model, the analysis of residuals of each model is carried out. Table 3 showed that we have two models that are very close in their AIC and BIC values. The models are ARIMA (3, 1,1) and ARIMA (3, 1,2). Therefore, further tests are necessary to determine the model. The residual test required that residuals are random with zero mean, constant variance and uncorrelated. Test for randomness of residuals are presented in Table 4.

Table 4: Comparison of different models (randomness test for residual)

Model	**ARIMA (3,1,1)**	**ARIMA (3,1,2)**
1 run above and below median	-58.474	-58.474
No. of runs above & below median	10	10
Expected no. of run	14.0	14.0
Large sample test statics	1.5011	1.5011
P – value	0.161177	0.161177
2- runs up and down		
No. of runs up & down	17	17
Expected no. of run	18.8887	18.8887
Large sample test statistics z	0.0898731	0.0898731
P – value	0.98716	0.98716
Test based on 1st 9 autocorrelation		
Large sample test Z	-0.04844	-0.04844
P – value	0.99984	0.99984

The randomness tests have identical results. Hence, the results are supportive of the randomness of residuals of both models at 95% significance level. Another feature of residuals is variance constant and not correlated. The ARCH –LM test for residuals is used. The ARCH-LM test reveals that ARIMA (3.1.1) better fits and describes the behavior of underlying series (Table 5).

Table 5: Comparison of arch – lm test

Model	F– Statistics	Probability
ARIMA (3,1,1)	0.1151	0.0975
ARIMA (3,1,2)	0.1247	0.9995

Estimation results

As the diagnostic checking tests showed (Tables 3 and 4), it is clear that MA model with lag1 more accurately forecasts Sokoto temperature. Therefore, ARIMA (3, 1, 1) model is selected. Table 6 presents the estimation results of ARIMA (3, 1, 1) model. According to the estimation results, the coefficient of MA (1) is significant at level 5% significance level.

Table 6: Estimation result of ARIMA (3, 1,1)

Variable	Coefficient	St. error	Probability
C	67.59076	49.9703	0.2010
MA(1)	-0.008812	0.20008	0.9692

R^2= 0.00078 C = 378.884 Statistic =0.0016 Prob. (F- Statistic) =0.09682221
Adj R^2 = 0.050 C = 381.581 D–Watson = 1.9811. The low coefficient of determination R^2 is not important due to differencing of the variable. The Durbin-Watson (DW) indicates no Serial correlation. The ARIMA (3, 1, 1) model can be rewritten in the form.

$$y_t = 44.3973 + 1.63821 y_{t-1} - 1.3764 y_{t-2} - 0.5891 y_{t-3} + 0.9861 \xi_{t-1} + \xi_t$$

Forecasting

ARIMA models are developed basically to forecast the corresponding variable. There are two kinds of forecasts: sample period forecasts and post sample period forecasts. The former are used to develop confidence in the model and the latter to generate genuine forecasts for use in planning and other purposes. The ARIMA model can be used to yield both these kinds of forecasts.
Forecasts for period of 14 days

95% Limits

Period	Forecast	Lower	Upper	Actual
1	25.1034	20.8267	29.380125.1034	
2	24.9196	19.9788	29.860524.9196	
3	24.8676	19.5518	30.183324.8676	
4	24.8531	19.2361	30.470124.8531	
5	24.8492	18.9590	30.739524.8492	
6	24.8485	18.7005	30.996424.8485	
7	24.8486	18.4542	31.242924.8486	
8	24.8489	18.2176	31.480224.8489	
9	24.8493	17.9892	31.709324.8493	
10	24.8497	17.7683	31.931124.8497	
11	24.8501	17.5542	32.146124.8501	
12	24.8506	17.3461	32.355024.8506	
13	24.8510	17.1437	32.558324.8510	
14	24.8514	16.9465	32.756424.8514	

Conclusion

The structure and behaviour of daily temperature of Sokoto town has been analyzed using ARIMA process. In doing so a much richer dynamic behavior of the series has been captured, achieved by integration of order one (1). A normality test performed on the mean and variance using the Anderson-Darling Normality Test at 95% confidence interval revealed that the mean is rightly skewed with a mean value of 28.799699 and variance of 18.003. The coefficient of skewness and kurtosis are -0.060749 and 2.994989 respectively. It is evident at 5% significant level that there are large swings in the data indicating non stationarity. Finally, it is modeled using the Box-Jenkins approach. The developed model that best fit the data was found to be ARIMA (3, 1, 1). From the forecast results obtained using the developed model, it can be seen that the forecasted temperature for two weeks showed decreasing trend for the first ten days, but started to increase towards the end. It can be observed that the rate of change in the daily temperature is very slow.

Recommendation

This research can be extended in several directions. For example, the seasonal structure may be examined. The possibility of segmented trends in this data is another issue that can be examined in future studies. Increasing global temperatures are causing a broad range of changes. Amounts and patterns of precipitation are changing. Changes in temperature and precipitation patterns increase the frequency, duration, and intensity of other extreme weather events, such as floods, droughts, heat waves, and tornadoes. Other effects of global warming include higher or lower agricultural yields, species extinctions, e.t.c. To this end, the results of this study with the knowledge of other factors should be used to come up with possible solutions to adverse temperature changes caused as a result of human activities.

References

Box, G. E. P, Jenkins, G. M. and Reinsel, G. C. (1994), "Time Series analysis, Forecasting and Control". 3rd Ed. Prentice Hall: Englewood Cliffs.

Dickey, D. A. and W. A. Fuller (1979), "Distribution of the estimators for Autoregressive Time Series with a Unit Root," Journal of the American Statistics Association, 74, p. 427-431.

Global warming (2013). Accessed from http://www.globalwarmingart.com/images/b/bb/1000_Year_Temperature_Comparison.png on 26/10/2013

Sokoto State Government Dairy (2002), "Ministry of Information, youth, sports and culture, Sokoto".

Time for Change (2003). "Accessed from http://timeforchange.org/cause-and-effect-for-global-warming" on 31/10/2013

Tsoho U.H, (2008), "Growth and History of the establishment of Makera Assada in Sokoto Metropolis to the year 2007". B.A project, History Department, Usmanu Danfodiyo University Sokoto.

Wikipedia, the free encyclopedia. Accessed from http://en.wikipedia.org/wiki/Sokoto on 26/10/2013