*Journal of the Nigerian Association of Mathematical Physics*
*Volume27*, (*July, 2014*), pp201 – 208


# Statistical Modelling and Prediction of Rainfall Time Series Data

*[1]T.O.Olatayo, [1]A.I. Taiwoand [2]R.A.Afolayan*

**[1]Department of Mathematical Sciences, Olabisi Onabanjo University, Ago-Iwoye, Ogun State, Nigeria.**
**[2]Department of Statistics, University of Ilorin, Ilorin Nigeria.**

## *Abstract*

***Climate and rainfall are highly non-linear and complicated phenomena, which require classical, modern and detailed models to obtain accurate prediction. In this paper, we present tools for modelling and predicting the behavioural pattern in rainfall phenomena based on past observations. The paper introduces three fundamentally different approaches for designing a model, the statistical method based on autoregressive integrated moving average (ARIMA), the emerging fuzzy time series (FTS) model and the non-parametric method(Theil's regression). In order to evaluate the prediction efficiency, we made use of 31 years of annual rainfall data from year 1982 to 2012 of Ibadan, Oyo State, Nigeria. The fuzzy time series model has it universe of discourse divided into 13 intervals and the interval with the largest number of rainfall data is divided into 4 sub-intervals of equal length. Three rules were used to determine if the forecast value under FTS is upward 0.75–point, middle or downward 0.25-point. ARIMA (1, 2, 1) was used to derive the weights and the regression coefficients, while the theil's regression was used to fit a linear model. The performance of the model was evaluated using mean squared forecast error (MAE), root mean square forecast error (RMSE) and Coefficient of determination ($R^2$). The study reveals that FTS model can be used as an appropriate forecasting tool to predict the rainfall, since it outperforms the ARIMA and Theil's models.***

**Keywords:** Fuzzy time series, Autoregressive integrated moving average, Theil's regression, Mean squared forecast error, Root mean square forecast error and Coefficient of determination

## 1.0 Introduction

Climate change seems to be the foremost global challenge facing humans at the moment, even though it seems that not all places on the globe are affected. World leaders, union leaders, pressure groups and others who have shown concern have been meeting to find a lasting solution to the 'acclaimed' dilemma. The scientific community has not been left out as causes and solutions are being proffered and it is expected to linger on for a long time, since rainfall is an indicator of climate change [1 – 3].

Rainfall is a climate parameter that affects the way and manner men lives. It affects every facet of the ecological system, flora and fauna inclusive. Hence, the study of rainfall is important and cannot be over emphasized [4]. Aside the beneficial aspect of rainfall, it can also be destructive in nature; natural disasters like floods and landslides are caused by rains [5]. Globally, lots of studies have been carried out on rainfall. A few of them is discussed briefly as follows. In [6], different trends across Sri Lanka using 100 years data were observed. Some parts recorded decreasing trend, some increasing trend while some locations showed no coherent trend. They also showed that the trend characteristics vary with the duration of the data analyzed. The trend analysis of rainfall over Jordan in [7] was examined picking three close-by locations. This study covered a period of 81 years (1922 – 2003). Although, different trends for different seasons across the three stations were observed, however, one of the stations showed a decline in both the rainy days and the total amount of rainfall after the mid 1950s. While in Turkey, the trend within a 64 year period (1929 – 1993) of rainfall for 96 stations was examined [8]. The overall result indicated that the trend in precipitation is downward, nonetheless, there are few stations that showed increasing trend.

Acknowledging some of the research that has been done, it is very important to discuss climatic changes as it has contributed to the instability of rainfall in Nigeria, then it becomes a very important and sensitive issue which requires adequate attention from governments, corporate organisations and researchers. Since climate and rainfall are highly non-linear and complicated

Corresponding author: T.O.OlatayoE-mail:otimtoy@yahoo.com,Tel.: +2348034052195

phenomena, which require serious and vivid investigation and analysis. Then, this research is centred on analysing the pattern and structure of rainfall over 30 years in Ibadan, South Western, Nigeria. Hence forecast values will be obtained in order to plan for the future.

In order to achieve our set objectives, classical, non-parametric and modern methods of discussing relationship and forecasting will be discussed. For classical forecasting method, we will consider autoregressive integrated moving average (ARIMA) which is a concept of autoregressive moving average while theil's regression will be used in the concept of non-parametric, where fuzzy time series method will be used in the concept of modern forecasting method. ARIMA is basically a linear statistical technique and has been quite popular for modelling the time series and rainfall forecasting due to ease in its development and implementation.

In contrast, fuzzy time series is another important modern forecasting method introduced by Song and Chissom in 1993 and it is believed that the theory of fuzzy time series overcome the drawback of the classical time series methods, it has the advantage of reducing the calculation time and simplifying the calculation process. Based on the theory of fuzzy time series, Song et al. presented some forecasting methods [9 – 12] and these methods are now being used in several fields to obtain meaningful results. Furthermore, theil's regression is a simple, non-parametric approach to fit a straight line to set of two points. This method was introduced by TheilSen in 1950 and it is has the ability to fit a linear trend when no assumptions about the population distribution from which the data was taken are known.

However, the three models will be used to forecast values for rainfall behaviour and the results will be compared to determine maybe the result obtained using classical forecasting method will better the result obtained for the non parametric and modern methods and vice verse.

## 2.0 Materials and Methods

### 2.1 Data Exploration

The pattern and general behaviour of the series is examined from the time plot. The series will be examined for stationarity, outliers and gaussianity. Test for stationarity will be carried out using correlogram. Details of the test procedures can be found in [13].

### 2.2 ARIMA Theory

*ARIMA* (autoregressive integrated moving average) models are generalizations of the simple *AR* model that use three tools for modelling the serial correlation in the disturbance. The first tool is the autoregressive, or *AR*, term. The $AR(1)$ model use only the first-order term, but in general, you may use additional, higher-order AR terms. Each AR term corresponds to the use of a lagged value of the residual in the forecasting equation for the unconditional residual. An autoregressive model of order $(p)$ , $AR(p)$ has the form:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t \qquad (1)$$

The second tool is the integration order term. Each integration order corresponds to differencing the series being forecast. A first-order integrated component means that the forecasting model is designed for the first difference of the original series. A second - order component corresponds to using second differences, and so on.

The third tool is the MA, or moving average term. A moving average forecasting model uses lagged values of the forecast error to improve the current forecast. A first order moving average term uses the most recent forecast error; a second-order term uses the forecast error from the two most recent periods, and so on. An MA(q ) has the form:

$$u_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \qquad (2)$$

The autoregressive and moving average specifications can be combined to form an ARMA (p, q) specification:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \qquad (3)$$

#### 2.2.1 Principles of ARIMA Modelling

In ARIMA forecasting, you assemble a complete forecasting model by using combinations of the three building blocks to be described below. The first step is forming an ARIMA model for a series of residuals by looking into its autocorrelation properties. We will make use the correlogram view of a series for this purpose. This phase of the ARIMA modelling procedure is called identification.

The next step is to decide what kind of ARIMA model to use. If the autocorrelation function dies off smoothly at a geometric rate, and the partial autocorrelations were zero after one lag, then a first-order autoregressive model is appropriate. Alternatively, if the autocorrelations were zero after one lag and the partial autocorrelations declined geometrically, a first order moving average process would seem appropriate.

#### 2.2.2 Estimating ARIMA Models

To specify your $ARIMA$ model, you will difference your dependent variable, if necessary, to account for the order of

integration and describe your structural regression model (dependent variables and regressors) and add any $AR\ and MA$ terms. The d operator can be used to specify differences of series. To specify first differencing, simply include the series name in parentheses after d. For example, $d(rainfall)$ specifies the first difference of rainfall.
More complicated forms of differencing may be specified with two optional parameters, $n\ and\ s, d(X,n)$ specifies the $nth$ order difference of the series $X$:

$$d(X,n) = (1-L)^n x \qquad (4)$$

Where $L$ is the lag operator.

## 2.3 Basic Concept of Fuzzy Time Series

The definition of fuzzy time series based on fuzzy sets in [14] and [15] as follows: Let $T$ be the universe of discourse, $T = \{t_1, t_2, \dots, t_n\}$ and let $Z$ be a fuzzy set in the universe of discourse $U$ defined as follows:

$$Z = f_z(t_1)/t_1 + f_z(t_2)/t_2 + \cdots + f_z(t_n)/t_n \qquad (5)$$

Where $f_z$is the membership function of $Z. f_Z: T \rightarrow [0,1], f_z(t_i)$ indicates the grade of membership of $t_i$in the fuzzy set $Z$, $f_z(t_i) \in [0,1]$ and $1 \leq i \leq n$.
Let $X(u)$ $(u = \cdots, 0,1,2, \dots)$ be the universe of discourse and be a subset of $R$, and let fuzzy set $f_i(u)$ $(i = 1.2, \dots)$ be defined in $X(u)$. Let $F(u)$ be a collection of $f_i(u)$ $(i = 1.2, \dots)$. Then, $F(u)$ is called a fuzzy time series of $X(u)$ $(u = \cdots, 0,1,2, \dots)$ .
If $F(u)$ is caused by $F(u-1)$, denoted by $F(u-1) \rightarrow F(u)$, then this relationship can be represented by $F(u) = F(u-1) \circ R(u, u-1)$, where the symbol $'' \circ ''$ denotes the Max-Min composition operator; $R(u, u-1)$ is a fuzzy relation between $F(u)$ and $F(u-1)$ and is called the first-order model of $F(u)$.
Let $F(u)$ be a fuzzy time series and let $R(u, u-1)$ be a first-order model of $(u)$ . If $R(u, u-1) = R(u-1, u-2)$ for any time $u$, then $F(u)$ is called a time-invariant fuzzy time series. If $R(u, u-1)$ is dependent on time $u$, that is, $R(u, u-1)$ may be different from $R(u-1, u-2)$ for any u , then $F(u)$ is called a time-variant fuzzy time series.

### 2.3.1 Fuzzy Time Series Model

Using the time-variant fuzzy time-series model, the following steps form the procedure.
**Step 1:** Define the universe of discourse within which fuzzy sets are defined.
**Step 2:** Partition the universe of discourse $T$ into several even and equal length intervals.
**Step 3:** Determine some linguistic values represented by fuzzy sets of the intervals of the universe of discourse.
**Step 4:** Fuzzify the rainfall data.
**Step 5:** Choose a suitable parameter $\omega$, where $\omega > 1$, calculate $R^{\omega}(u, u-1)$ and forecast the rainfall as follows:

$$F(u) = F(u-1) \circ R^{\omega}(u, u-1) \qquad (6)$$

where$F(u)$ denotes the forecasted fuzzy rainfall of year $u$, $F(u-1)$denotes the fuzzified rainfall of year $u-1$, and

$$R^{\omega}(u, u-1) = F^T(u-2) \times F(u-1) \cup F^T(u-1) \times F(u-2) \cup \dots \cup F^T(u-\omega) \times F(u-\omega+1) \qquad (7)$$

where$\omega$ is called the "model basis" denoting the number of years before $u$, $'' \times ''$ is the Cartesian product operator, and $T$ is the transpose operator.
**Step 6:** Defuzzify the forecasted fuzzy rainfall using neural nets.
It very important to note that we will divide each interval derived in $step$ 2 into four subintervals of equal length, where the 0.25-point and 0.75-point of each interval are used as the upward and downward forecasting points of the forecasting. Three rules were used and they are:
(1). If |(the difference of the rainfall between years $n-2$ and $n-1$)|/2 $>$ half of the length of the interval corresponding to the fuzzified rainfall $A_j$ with the membership value equal to 1, then the trend of the forecasting of this interval will be upward and the forecasting rainfall falls at the 0.75-point of this interval; if |(the difference of the rainfall data between years $n-2$ and $n-1$)|/2 $=$ half of the length of the interval corresponding to the fuzzified rainfall $A_j$ with the membership value equal to 1, then the forecasting rainfall falls at the middle value of this interval; if |(the difference of the rainfall data between years $n-2$ and $n-1$) )|/2 $<$ half of the length of the interval corresponding to the fuzzified rainfall $A_j$ with the membership value equal to 1, then the trend of the forecasting of this interval will be downward, and the forecasting rainfall falls at the 0.25-point of the interval.
(2). If (|the difference of the differences between years n-1 and n-2 and between years n-2 and n-3| × 2 $+$ the rainfall data of year n-1) or (the rainfall of year n-1 - |the difference of the differences between years n-1 and n-2 and between years n-2 and n-3| × 2) falls in the interval corresponding to the fuzzified rainfall $A_j$ with the membership value equal to 1, then the trend of the forecasting of this interval will be upward, and the forecasting rainfall falls at the 0.75-point of the interval of the corresponding fuzzified rainfall $A_j$ with the membership value equal to 1; if (|the difference of the differences between years

n-1 and n-2 and between years n-2 and n-3|/2＋ the rainfalls of year n-1) or (the rainfalls of year n-1 - |the difference of the differences between years n-1 and n-2 and between years n-2 and n-3|/2) falls in the interval of the corresponding fuzzified rainfall $A_j$ with the membership value equal to 1, then the trend of the forecasting of this interval will be downward, and the forecasting value falls at the 0.25-point of the interval of the corresponding fuzzified rainfall $A_j$ with the membership value equal to 1; if neither is the case, then we let the forecasting rainfall be the middle value of the interval corresponding to the fuzzified rainfall $A_j$ with the membership value equal to 1.

(3). If (|the difference of the differences between years n-1 and n-2 and between years n-2 and n-3|/2 ＋ the rainfall data of year n-1) or (the rainfall data of year n-1 - |the difference of the differences between years n-1 and n-2 and between years n-2 and n-3|/2) falls in the interval of the corresponding fuzzified rainfall $A_j$ with the membership value equal to 1, then the trend of the forecasting of this interval will be downward, and the forecasting rainfall falls at the 0.25-point of the interval corresponding to the fuzzified rainfall $A_j$ with the membership value equal to 1; if (|the difference of the differences between years n-1 and n-2 and between years n-2 and n-3| × 2 ＋ the rainfall data of year n-1) or (the rainfall data of year n-1 - |the difference of the differences between years n-1 and n-2 and between years n-2 and n-3| × 2) falls in the interval corresponding to the fuzzified rainfall $A_j$ with the membership value equal to 1, then the trend of the forecasting of this interval will be upward, and the forecasting rainfall falls at the 0.75-point of the interval corresponding to the fuzzified rainfall $A_j$ with the membership value equal to 1; if neither is the case, then we let the forecasting rainfall be the middle value of the interval corresponding to the fuzzified rainfall $A_j$ with the membership value equal to 1.

## 2.4 Theil's Regression

This is a simple and non-parametric approach for fitting a straight line to a set of $(x, y)$-points is the theil's method which assumes that points $(x_1, y_1)$, $(x_2, y_2), \dots, (x_N, y_N)$ are described by the equation; $y = a + bX$.
The calculation of $a$ and $b$ follows the steps outlined below;

- All $N$ data points are ranked in ascending order of $X$ values.
- The data are separated into two equal $(m)$ groups, the low $(L)$ and the high $(H)$ group. If $N$ is odd the middle data point is not included in either group.
- The slope $b_i$ is calculated for all points of each group, i.e, $b_i = (y_{H,I} - y_{L,.i})/(X_{H,I} - X_{I.i})$ for $i = 1,2, \dots, m$. (8)
- The median of the $m$ slope values $b_1, b_2, \dots, b_m$ is calculated and it is taken as the best estimate of the slope $(b)$ of the line, i.e $b = median(b_1, b_2, \dots, b_m)$
- For each data point $(x_i, y_i)$, the value of the intercept $a_i$ is calculated using the previously calculated slope $b$, that is $a_i = y_i - bX_i$ for $i = 1,2, \dots, N$. The median of the $N$ intercept values $a_1, a_2, \dots, a_N$ is calculated using and it is taken as the best estimate of the intercept $(a)$ of the line, that is $a = median(a_1, a_2, \dots, a_N)$.

## 2.5 Forecast Evaluation

Forecasts of ARIMA, Fuzzy Time series and Theil's regression will be computed for in-sample values. The optimal forecasts values are then evaluated using the mean squared forecast error (MAE) defined as,

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left(\hat{X}_t - X_t\right)^2 \qquad (9)$$

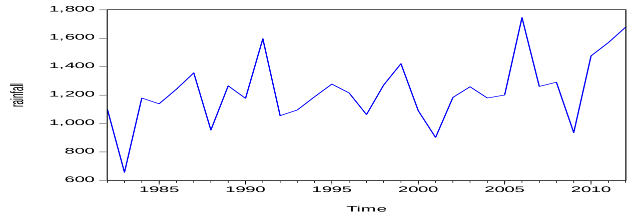
the root mean square forecast error (RMSE) is defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\hat{X}_t - X_t\right)^2} \qquad (10)$$

The actual and predicted values for corresponding $t$ values are denoted by $\hat{X}_t$ *and* $X_t$respectively. The smaller the values of *RMSE* and *MAE*, the better the forecasting performance of the model.

## 3.0 Results and Discussions.

The annual rainfall of Ibadan in South Western region of Nigeria which is bounded by$3^0 53', 7^0 22'$ will be used for this study. The data was obtained from the Nigerian Meteorological Agency, Lagos. It consists of the annual rainfall from 1981 to 2012 (31 years).



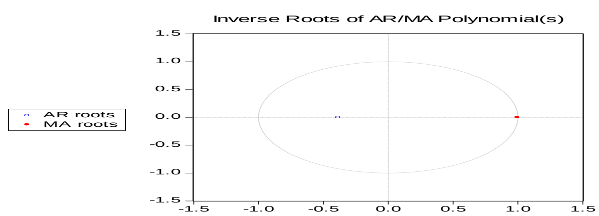

$Fig.1: Tlme\ Plot\ of\ Rainfall\ data\ in\ Ibadan\ from\ 1982-2012$

***Table 1:*** *Unit Root Test Using Augmented Dickey-Fuller (ADF)*

| | | | t-Statistic | Prob.* |
|---|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | | -3.584924 | 0.0150 |
| Test critical values: | 1% level | | -3.769597 | |
| | 5% level | | -3.004861 | |
| | 10% level | | -2.642242 | |
| *MacKinnon (1996) one-sided p-values. | | | | |

$\boldsymbol{Table\ 2}: Correlogram\ of\ D(Rainfall, 2)$

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| *****\| . \| | *****\| . \| | 1 | -0.634 | -0.634 | 12.896 | 0.000 |
| . \|**. \| | .**\| . \| | 2 | 0.249 | -0.256 | 14.956 | 0.001 |
| .**\| . \| | ***\| . \| | 3 | -0.263 | -0.407 | 17.351 | 0.001 |
| . \|**. \| | .**\| . \| | 4 | 0.231 | -0.257 | 19.263 | 0.001 |
| . \| . \| | . \|* . \| | 5 | 0.019 | 0.126 | 19.276 | 0.002 |
| .**\| . \| | . *\| . \| | 6 | -0.222 | -0.200 | 21.208 | 0.002 |
| . \|* . \| | . *\| . \| | 7 | 0.185 | -0.096 | 22.606 | 0.002 |
| . \| . \| | . \|* . \| | 8 | -0.050 | 0.102 | 22.712 | 0.004 |
| . \| . \| | . \| . \| | 9 | 0.050 | 0.002 | 22.825 | 0.007 |
| . *\| . \| | . \| . \| | 10 | -0.110 | -0.018 | 23.398 | 0.009 |
| . \| . \| | . \| . \| | 11 | 0.027 | -0.040 | 23.435 | 0.015 |
| . \|* . \| | . \| . \| | 12 | 0.076 | -0.060 | 23.744 | 0.022 |




$Fig. 2\ Inverse\ Root\ of\ ARMA$

$\boldsymbol{Table\ 3}.\ Correlogram\ of\ Residuals$

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| . *\| . \| | . *\| . \| | 1 | -0.117 | -0.117 | 0.4240 | |
| .**\| . \| | .**\| . \| | 2 | -0.290 | -0.308 | 3.1417 | |
| .*\| . \| | *\| . \| | 3 | -0.276 | -0.399 | 5.7067 | 0.017 |
| . \|** \| | . \|*. \| | 4 | 0.010 | 0.241 | 6.602 | 0.003 |
| . \| . \| | . *\| . \| | 5 | -0.034 | -0.151 | 6.645 | 0.009 |
| .**\| . \| | .**\| . \| | 6 | -0.265 | -0.280 | 7.318 | 0.006 |
| . \| . \| | . \|**. \| | 7 | 0.068 | 0.239 | 8.504 | 0.013 |
| . \|**. \| | . \|* . \| | 8 | 0.341 | 0.166 | 8.679 | 0.004 |
| . \| . \| | . \| . \| | 9 | -0.007 | -0.010 | 8.781 | 0.007 |
| .**\| . \| | . \|* . \| | 10 | -0.009 | 0.080 | 9.560 | 0.003 |
| . \| . \| | . \| . \| | 11 | -0.043 | -0.018 | 9.651 | 0.005 |
| . \|* . \| | . *\| . \| | 12 | 0.015 | -0.074 | 10.643 | 0.004 |

**Fuzzy Time Series Steps**

$Step$ 1 : The universe of discourse $U = [600, 1800]$ and it is partitioned into six even and equal length intervals $u_1, u_2, u_3, u_4, u_5$ and $u_6$ where $u_1 = [600, 800]$, $u_2 = [800, 1000]$, $u_3 = [1000, 1200]$, $u_4 = [1200, 1400]$, $u_5 = [1400, 1600]$, $u_6 = [1600, 1800]$,

$Step$ 2: Get a statistics of the distribution of the rainfall data in each interval.

**Table 4.** The distribution of the historical rainfall data

| *Interval* | $[600, 800]$ | $[800, 1000]$ | $[1000, 1200]$ | $[1200, 1400]$ | $[1400, 1600]$ | $[1600, 1800]$ |
|---|---|---|---|---|---|---|
| *Number of rainfall data* | 1 | 3 | 11 | 10 | 4 | 2 |

The universe of discourse $[600, 1800]$ is re-divided into the following intervals:
$u_{1,1} = [600, 700], u_{1,2} = [700, 800], u_2 = [800, 1000], u_{3,1} = [1000, 1050]$, $u_{3,2} = [1050, 1100$ , $u_{3,3} = [1100,1150]$ , $u_{4,1} = [1200, 1266, u_{4,2} = [1266, 1334]$,
$u_{4,3} = [1334, 1400], u_{5,1} = [1400, 1500]$ ,
$u_{5,2} = [1500, 1600]$ and $u_6 = [1600, 1800]$

$Step$ 3**:** We define each fuzzy set $A_i$ based on the re-divided intervals and fuzzify the rainfall data, where fuzzy set $A_i$denotes a linguistic value of the rainfall data represented by a fuzzy set and $1 \le i \le 13$. The membership values of fuzzy set $A_i$ either are $0, 0.5$ $or$ $1$. Then, we fuzzify the rainfall data and the linguistic values of the rainfall$A_1, A_2, \dots , A_{13}$. The reason for fuzzifying the rainfall data into fuzzified rainfall is to translate crisp values into fuzzy sets to get a fuzzy time series.

$Step$ 4: Establishing fuzzy logical relationships based on the fuzzified rainfall:

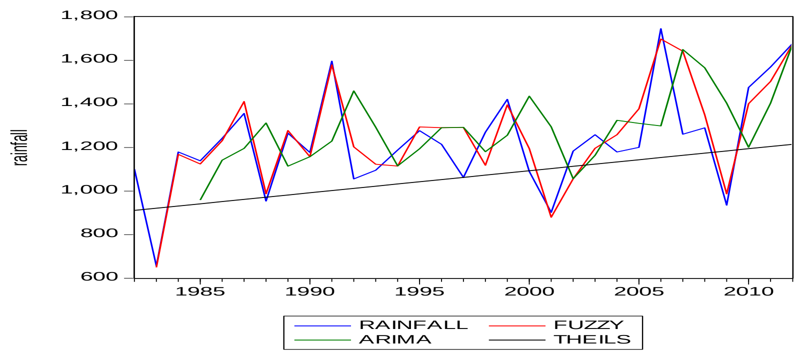
$$A_j \to A_q$$
$$A_j \to A_r$$

where the fuzzy logical relationship $''A_j \to A_q''$ denotes " if the fuzzified rainfall data of year $n - 1$ is $A_j$ then the fuzzified rainfall of year $n$ is $A_q$".

$Step$ 5:Divide each interval derived in $step$ 2 into four subintervals of equal length, where the 0.25-point and 0.75-point of each interval are used as the upward and downward forecasting points of the forecasting.

**Table 5.** Forecast Result of Fuzzy Time Series

| Year | Rainfall | Trend of the Forecasting | Forecasting |
|---|---|---|---|
| 1982 | 1100.8 | | |
| 1983 | 656.2 | Middle value | 649.5 |
| 1984 | 1179.3 | Upward; 0.75 - point | 1168.5 |
| 1985 | 1138.7 | Downward; 0.25 - point | 1124.5 |
| 1986 | 1242 | Downward; 0.25 - point | 1231 |
| 1987 | 1356.8 | Upward; 0.75 - point | 1411 |
| 1988 | 954 | Upward; 0.75 - point | 987.5 |
| 1989 | 1265.5 | Middle value | 1278 |
| 1990 | 1177.2 | Upward; 0.75 - point | 1157.5 |
| 1991 | 1596.4 | Upward; 0.75 - point | 1579.5 |
| 1992 | 1055.5 | Middle value | 1203 |
| 1993 | 1095.5 | Middle value | 1123.25 |
| 1994 | 1188.2 | Upward; 0.75 - point | 1115 |
| 1995 | 1277.5 | Middle value | 1294.5 |
| 1996 | 1214.5 | Downward; 0.25 - point | 1291 |
| 1997 | 1062.9 | Middle value | 1292 |
| 1998 | 1270.7 | Middle value | 1118.5 |
| 1999 | 1421.5 | Downward; 0.25 - point | 1396.25 |
| 2000 | 1090.3 | Middle value | 1195.25 |
| 2001 | 901.7 | Middle value | 879.5 |
| 2002 | 1183.8 | Upward; 0.75 - point | 1056.25 |
| 2003 | 1258.9 | Middle value | 1196.5 |
| 2004 | 1179.3 | Upward; 0.75 - point | 1258.5 |
| 2005 | 1200.7 | Downward; 0.25 - point | 1378 |
| 2006 | 1745.8 | Middle value | 1698 |
| 2007 | 1261.2 | Middle value | 1642.5 |
| 2008 | 1290.6 | Middle value | 1350.5 |
| 2009 | 935.5 | Middle value | 987.25 |
| 2010 | 1475.8 | Middle value | 1401.5 |
| 2011 | 1569.5 | Upward; 0.75 - point | 1503.25 |
| 2012 | 1678.2 | Upward; 0.75 - point | 1675.5 |




$Fig. 3$ *Actual and Forecast Raifall data from* $1982 - 2012$

***Table 6.*** *Mean Absoloute Errors and Root Mean Square Error Values*

| Model | MAE | RMSE | $R^2$ |
|---|---|---|---|
| ARIMA | 110.23 | 10.49 | 0.97882 |
| Fuzzy Time Series | 85.45 | 9.24 | 0.98456 |
| Theil's Regression | 226.12 | 15.03 | 0.83346 |

## 3.1 Discussion

It is evidence from the time plots that rainfall data displays series of cyclical behaviour and this is due to seasonal changes yearly. For autoregressive integrated moving average, model building commenced with the examination of the plot of the series, the sample plot of the autocorrelation (ACF) and partial autocorrelation (PACF) model description. The time plot of the original series ($Fig. 1$) shows stationarity as confirmed by the Augmented Dickey-fuller test in *Table 1* with a p-value of 0.05, but with seasonal trend.

Since the order of integration of the differenced rainfall series in *Table 2* is two, then $d = 2$ *and* a close look of the ACF and PACF of the differenced data in *Table 2* revealed the ACF dies off smoothly at a geometric rate and the partial autocorrelations were zero after one lag and the autocorrelations were zero after one lag and the partial autocorrelations declined geometrically, these behaviour shows that $ARIMA$ (1,2,1) is the appropriate model for the differenced rainfall series, that is $(1 - \rho_1 L)\Delta^2 u_t = (1 - \theta L)\varepsilon_t$. Therefore the fitted model is given as:

$$y_t = 4.37 + u_t$$

$$(1 - 0.39L)u_t = (1 - 0.99L)\varepsilon_t$$

With the white noise variance $\hat{\sigma}_\varepsilon^2$ estimated as 17452. In order to use the model obtained for forecast some model diagnostic test were carried out. The inverse root of $ARMA$ in *Fig. 2* shows that the estimated ARMA process is (covariance) stationary, since all AR roots lie *inside* the unit circle and the estimated ARMA process is invertible, since all MA roots should lie *inside* the unit circle. The correlogram has no significant spike and all subsequent $Q$-statistics are not highly significant. This result clearly indicates there is no need for respecification of the model. However, the forecast of the yearly rainfall from 1982 to 2012 deviated slightly from the original data, *see* $Fig. 3$.

Under fuzzy time series, we made use of the visual Basic Version 6.0 on a Pentium 4 PC. *Table 4* summarizes the forecasting results of fuzzy time series method from 1982 to 2012, where the universe of discourse is divided into 13 intervals and the interval with the largest number of rainfall data is divided into 4 sub-intervals of equal length. The fuzzy time series forecast of the yearly rainfall data from 1982 to 2012 did not deviated much from the original data, *see Table* 5 *and Fig.* 3.
Using the non-parametric method (theil's regression), we obtain a fitted linear model: $Y = 900.98 + 10.12(X)$, where $Y$ represents rainfall data and $X$ represents time.

## 3.2 A Comparison of Different Forecasting Methods

The performance measures of ARIMA, FTS and theil's regression models in terms of numerical computations are shown in Table 6. The table indicates that the FTS model outperforms both the ARIMA and theil's regression model. While the ARIMA model is better than the theil's regression model. The MAE for ARIMA model and theil's regression are 110.23 and 226.12 respectively. While the same MAE is considerably lower at 85.45 for FTS model. The other performance measures such as RMSE and $R^2$ depict that the FTS forecast is superior to ARIMA and theil's regression forecast. The forecast graph in fig. 3 as well shows clearly that FTS forecast did not deviate much from the original data compared to the two other models. Therefore, our study establishes that FTS method should be favoured as an appropriate forecasting tool to model and predict annual rainfall in Ibadan South Western, Nigeria.

## 4.0 Conclusion

Complexity of the nature of annual rainfall record has been studied using FTS, ARIMA and Theil's regression techniques. An annual rainfall data spanning over a period of 1982 – 2012 of Ibadan in South Western, Nigeria was used to develop and test the models. The study reveals that FTS model can be used as an appropriate forecasting tool to predict the rainfall series, which performs better than the ARIMA and Theil's regression models.

## 5.0 Acknowledgement

We want to acknowledge the effort of Nigerian Metrological Agency Lagos, Nigeria; for make available the Data that was used in this analysis of this research work.

## References

[1] Adger, W.N, Hug S, Brown K, Conway D. and Hulme .M (2003). Adaptation to climatechange in the developing world. Proc. Dev. Stud., 3(3): 179 - 195.

[2] Frich P, Alexander L.V, Della-Marta P, Gleason B, Haylock M, Klein Tank A.M.G andPeterson .T (2002). Observed coherent changes in climatic extremes during the secondhalf of twentieth century. Clim. Res., 19: 193 - 212.

[3] Novotny, E.VandSfefan H.G (2007). Stream flow in Minnesota: indicator of climateChange, J. Hydro., 334: 319-333.

[4] Obot, N.I and Onyeukwu, N.O (2010*)*.Trend of rainfall in Abeokuta, Ogun State, Nigeria: A 2-year experience (2006-2007). J. Env. Iss. Agric. Dev. Count., 2(1): 70 - 81.

[5] Ratnayake U and Herath .S (2005). Changing rainfall and its impacts on landslides in Sri Lanka. J. Mou. Sci., 2(3): 218-224.

[6] Jayawardene, K.H.W.I (2005). Trends of rainfall in Sri Lanka over the last century, J. of Physics, vol. 6, pp. 7 – 17.

[7] Smadi M.M and Zghoul A (2006). A sudden change in rainfall characteristics in Amman, Jordan during the Mid 1950s. Am. J. Env. Sci., 2(3): 84 - 91.

[8] Partal .T and Kahya .E (2006). Trend analysis in Turkish precipitation data. Hydrol. Proc., 20: 2011 – 2026.

[9] Song, Q. (2003). A note on fuzzy time series model selection with sampleAutocorrelation functions. *Cybernetics and Systems: An International Journal*, 34:93:107.

[10] Song, Q. and Chissom, B. S. (1993). Fuzzy time series and its models, *FuzzySets and Systems*, 54: 269-277.

[11] Song, Q. and Chissom, B. S. (1993). Forecasting enrollments with fuzzy time series –Part I. *Fuzzy Sets and Systems*, 54: 1-9.

[12] Song, Q. and Leland, R. P. (1996). Adaptive learning defuzzification techniques andapplications. *Fuzzy Sets and Systems*, 81: 321-329.

[13] Box, G.E.P. and Jenkins, G.M., 1976. Time Series Analysis, Forecasting and ControlHolden Day, CA, San Francisco.

[14] Song, Q. and Chissom, B. S. (1994). Forecasting enrollments with fuzzy time series –Part II. *Fuzzy Sets and Systems*, 62: 1-8

[15] Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8: 338 – 353.

[16] Nigerian Meteorological Agency (2012) Lagos, Nigeria.

[17] TheilSen (1950), Definition and Methodology of theil's regression