Global Journals LaTeX JournalKaleidoscope™

Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

Statistical Modelling and Prediction of Rainfall Time Series Data

A.I. Taiwo[1] and A,I,TAIWO[2]

[1] Olabisi Onabanjo University, Ago-Iwoye,Ogun State,Nigeria

Received: 6 December 2013 Accepted: 4 January 2014 Published: 15 January 2014

Abstract

Climate and rainfall are highly non-linear and complicated phenomena, which require classical, modern and detailed models to obtain accurate prediction. In order to attain precise forecast, a modern method termed fuzzy time series that belongs to the first order and time-variant method was used to analyse rainfall since it has become an attractive alternative to traditional and non-parametric statistical methods. 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(FST) 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 South West, 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 FST is upward 0.75â??"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 (. 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.

Index terms— fuzzy time series, autoregressive integrated moving average, theil?s regression, mean squared forecast error, root mean square forecast error and co

1 Introduction

limate 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. One of the indicators of climate change is rainfall (Adger et al., 2003;Frich et al., 2002; **??**ovotny and Stefan, 2007).

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 (Obot and Onyeukwu, 2010). Aside the beneficial aspect of rainfall, it can also be destructive in nature; natural disasters like floods and landslides are caused by rain (Ratnayake and Herath, 2005).

Globally, lots of studies have been carried out on rainfall. A few of them is discussed briefly; Jayawardene et al. **??**2005) observed different trends across Sri Lanka using 100 years data. 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. Smadi and Zghoul (2006) examined the trend analysis of rainfall over Jordan picking three close-by locations. Their study covered a period of 81 years . 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, Partal and Kahya (2006) examined the trend within a 64 year period of rainfall for 96 stations. 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 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 South West, 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 modeling 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 [Song (2003); and Song and Leland (1996)] 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 Theil Sen in 1950 and it is has the ability to fit a linear trend when no assumptions about the population distribution from which the data 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 II.

3 Theory and Methods

4 a) 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 Box and Jenkins (1976).

5 b) ARIMA Theory

ARIMA (autoregressive integrated moving average) models are generalizations of the simple AR model that use three tools for modeling the serial correlation in the disturbance. The first tool is the autoregressive, or AR, term. The ????(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 (??) , ????(??) has the form:?? ?? = ?? 1 ?? ???1 + ?? 2 ?? ???2 + ? + ?? ?? ?? ????? + ?? ??

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:?? ?? = ?? ?? + ?? 1 ?? ???1 + ?? 2 ?? ???2 + ? + ?? ?? ?? ?????

The autoregressive and moving average specifications can be combined to form an ARMA (p, q) specification:?? ?? = ?? 1 ?? ???1 + ?? 2 ?? ???2 + ? + ?? ?? ?? ????? + ?? ?? + ?? 1 ?? ???1 + ?? 2 ?? ???2 + ? + ?? ?? ?? ????? i. Principles of ARIMA Modeling

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 modeling 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 firstorder 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.

6 ii. Estimating ARIMA Models

To specify your ?????????? 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 ???? ?????????? 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, ??(????????????????) specifies the first difference of rainfall.

More complicated forms of differencing may be specified with two optional parameters, ?? ?????? ??, ??(??, ??) specifies the ????? order difference of the series ?? :??(??, ??) = (1 ? ??) ?? ??

Where ?? is the lag operator.

(

(3) c) Basic Concept of Fuzzy Time Series 1994) proposed the definition of fuzzy time series based on fuzzy sets in Zadeh (1965) as follows: Let ?? be the universe of discourse, ?? = {?? 1 , ?? 2 , ? , ?? ?? } and let ?? be a fuzzy set in the universe of discourse ?? defined as follows:?? = ?? ?? (?? 1) ?? 1 ? + ?? ?? (?? 2) ?? 2 ? + ? + ?? ?? (?? ??) ?? ?? ? (5)

where ?? ?? is the membership function of ??. ?? ?? : ?? ? [0,1], ?? ?? (?? ??) indicates the grade of membership of ?? ?? in the fuzzy set ??, ?? ?? (?? ??) ? [0,1] and 1 ? ?? ? ??.

Let ??(??) (?? = ? , 0,1,2, ?) be the universe of discourse and be a subset of ??, and let fuzzy set ?? ?? (??) (?? = 1.2, ?) be defined in ??(??). Let ??(??) be a collection of ?? ?? (??) (?? = 1.2, ?). Then, ??(??) is called a fuzzy time series of ??(??) (?? = ? , 0,1,2, ?) .

If ??(??) is caused by ??(?? ? 1), denoted by ??(?? ? 1) ? ??(??), then this relationship can be represented by ??(??) = ??(?? ? 1) ? ??(??, ?? ? 1), where the symbol ?? ? ?? denotes the Max-Min composition operator; ??(??, ?? ? 1) is a fuzzy relation between ??(??) and ??(?? ? 1) and is called the first-order model of ??(??).

Let ??(??) be a fuzzy time series and let ??(??, ?? ?

7 i. 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 ?? 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 ð ??”ð ??”, where ð ??”ð ??” > 1, calculate ?? ð ??”ð ??” (??, ?? ? 1) and forecast the rainfall as follows:??(??) = ??(?? ? 1) ? ?? ð ??”ð ??” (??, ?? ? 1)

where ??(??) denotes the forecasted fuzzy rainfall of year ??, ??(?? ? 1) denotes the fuzzified rainfall of year ?? ? 1, and?? ð ??”ð ??” (??, ?? ? 1) = ?? ?? (?? ? 2) × ??(?? ? 1) ? ?? ?? (?? ? 1) × ??(?? ? 2) ? ? ? ?? ?? (?? ? ð ??”ð ??”) × ??(?? ? ð ??”ð ??” + 1)

where ð ??”ð ??” is called the ”model basis” denoting the number of years before ??, ?? × ?? is the Cartesian product operator, and ?? 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 ???????? 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:

interval corresponding to the fuzzified rainfall ?? ?? with the membership value equal to 1.

8 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 ?? ?? 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 ?? ?? 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 ?? ?? 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 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 ?? ?? with the membership value equal to 1.

9 d) Theil's Regression

This is a simple and non-parametric approach for fitting a straight line to a set of -points is the theil's e) 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,?????? = 1 ?? ???? ? ?? ? ?? ?? ? 2 ?? ??=1**(8)** the root mean square forecast error (RMSE) is define as???????? = ? 1 ?? ???? ? ?? ? ?? ?? ? 2 ?? ??=1(**9**)

The actual and predicted values for corresponding ?? values are denoted by ?? ? ?? ?????? ?? ?? respectively. The smaller the values of RMSE and MAE, the better the forecasting performance of the model.

10 f) Data

The annual rainfall of Ibadan in South West 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). The universe of discourse [600, 1800] is redivided into the following intervals:

11 Fuzzy Time Series Steps

12 Results and 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 (??????. 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 (fig. **??**) is two, then ?? = 2 ?????? a close look of the ACF and PACF of the differenced data in (fig. **??**) 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 ?????????? (1,2,1) is the appropriate model for the differenced rainfall series, that is (1 ? ?? 1 ??)? 2 ?? ?? = (1 ? ????)?? ?? .Therefore the fitted model is given as:?? ?? = 4.37 + ?? ?? (1 ? 0.39??)?? ?? = (1 ? 0.99??)?? ??

With the white noise variance ?? ? ?? 2 estimated as 17452. In order to use the model obtained for forecast some model diagnostic test were carried out. The inverse root of ???????? in (fig. 3) 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, ?????? ??????. (5).

Under fuzzy time series, we made use of the visual Basic Version 6.0 on a Pentium 4 PC. Tab. 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, ?????? ??????. (4) ?????? ??????. (5).

Using the non-parametric method (theil's regression), we obtain a fitted linear model: ?? = 900.98 + 10.12(??), where ?? represents rainfall data and ?? represents time.

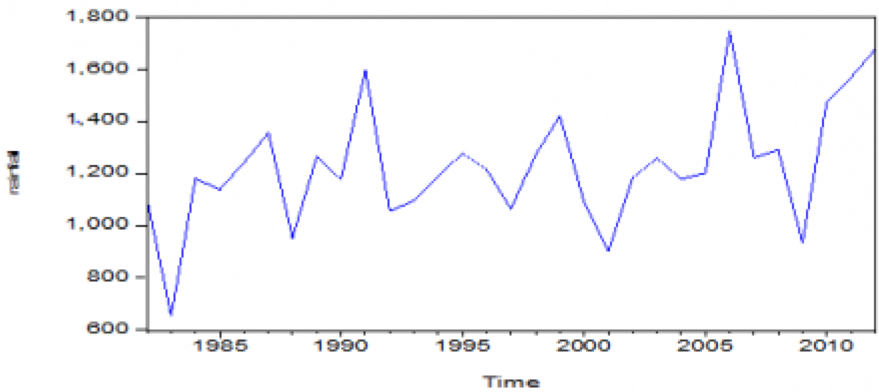
13 a) A Comparison of Different Forecasting Methods

The performance measures of ARIMA, FTS and theil's regression models in terms of numerical computations are shown in 226.12 respectively. While the same MAE is considerably lower at 85.45 for FTS model. The other performance measures such as RMSE and ?? 2 depict that the FTS forecast is superior to ARIMA and theil's regression forecast. The forecast graph in fig. 5 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 FST method should be favoured as an appropriate forecasting tool to model and predict annual rainfall.

14 V. Conclusion

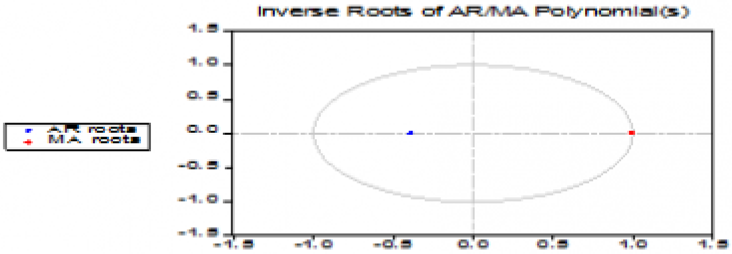
Complexity of the nature of annual rainfall record has been studied using FST, ARIMA and Theil's regression techniques. An annual rainfall data spanning over a period of 1982 -2012 of Ibadan in South West, Nigeria was used to develop and test the models. The study reveals that FST model can be used as an appropriate forecasting tool to predict the rainfall, which out performs the ARIMA and Theil's regression model. [1]

Figure 1: Statistical



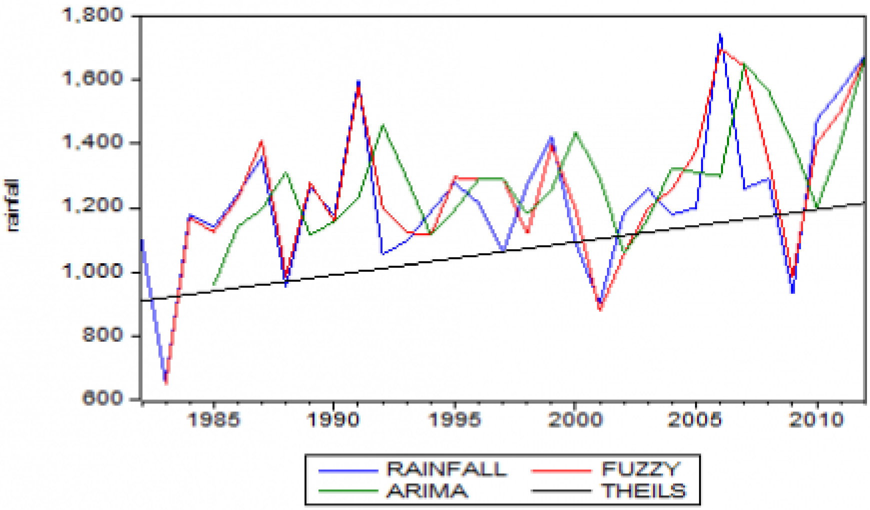
1

Figure 2: 1 .



3

Figure 3: Figure 3 :



21

Figure 4: Figure 2 :Figure 1 :

Year 2014
4
Volume XIV Issue I Version I
D D D D) G
(
Global Journal of Computer Science and Technology

method (?? which assumes that points

[Note: 1 , ?? 1), (?? 2 , ?? 2), ? , (?? ?? , ?? ??) are described by the equation; ?? = ?? + ????.The calculation of ?? and ?? follows the steps outlined below;? All ?? data points are ranked in ascending order of ?? values. ?]

Figure 5:

1

t-Statistic	Prob.*

[Note: Figure 3 : Inverse Root of ARMA]

Figure 6: Table 1 :

2

Number of rainfall data	1	3	11	10	4	2

Figure 7: Table 2 :

3

Year	Rainfall	Trend of the Forecasting	Forecasting Year 2014 D D D D D D D D) G (

Figure 8: Table 3 :

4

Model	MAE	RMSE	?? ??
ARIMA	110.23	10.49	0.97882
Fuzzy Time Series	85.45	9.24	0.98456
Theil's Regression IV.	226.12	15.03	0.83346

Figure 9: Table 4 :

4

Figure 10: Table 4 .

[Mou. Sci] , *Mou. Sci* 2 (3) p. .

[Zadeh ()] , L A Zadeh . *Fuzzy sets. Information and Control* 1965. 8 p. .

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

[Song ()] 'A note on fuzzy time series model selection with sample autocorrelation functions'. Q Song . *Cybernetics and Systems: An International Journal* 2003. 34 p. .

[Smadi and Zghoul ()] 'A sudden change in rainfall characteristics in Amman, Jordan during the Mid 1950s'. M M Smadi , A Zghoul . *Am. J. Env. Sci* 2006. 2 (3) p. .

[Adger et al. ()] 'Adaptation to climate change in the developing world'. W N Adger , S Hug , K Brown , D Conway , M Hulme . *Proc. Dev. Stud* 2003. 3 (3) p. .

[Song and Leland ()] *Adaptive learning defuzzification techniques and applications. Fuzzy Sets and Systems*, Q Song , R P Leland . 1996. 81 p. .

[Ratnayake and Herath ()] *Changing rainfall and its impacts on landslides in Sri Lanka*, U Ratnayake , S Herath . 2005.

[Sen ()] *Definition and Methodology of theil's regression*, Theil Sen . 1950.

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

[Song and Chissom ()] 'Forecasting enrollments with fuzzy time series -Part II'. Q Song , B S Chissom . *Fuzzy Sets and Systems* 1994. 62 p. .

[Song and Chissom ()] *Fuzzy time series and its models. Fuzzy Sets and Systems*, Q Song , B S Chissom . 1993. 54 p. .

[Intercomparison of conceptual models used in operational hydrological forecasting World Meteorological Organization ()] 'Intercomparison of conceptual models used in operational hydrological forecasting'. *World Meteorological Organization* 1975. (429) . (Technical report) (World Meteorological Organization)

[Frich et al. ()] 'Observed coherent changes in climatic extremes during the second half of twentieth century'. P Frich , L V Alexander , P Della-Marta , B Gleason , M Haylock , Klein Tank , Amg Peterson , T . *Clim. Res* 2002. 19 p. .

[References Références Referencias] *References Références Referencias*,

[Novotny and Sfefan ()] 'Stream flow in Minnesota: indicator of climate change'. E V Novotny , H G Sfefan . *J. Hydro* 2007. 334 p. .

[Box and Jenkins ()] *Time Series Analysis, Forecasting and Control*, G E P Box , G M Jenkins . 1976. Holden Day, CA, San Francisco.

[Partal and Kahya ()] 'Trend analysis in Turkish precipitation data'. T Partal , E Kahya . *Hydrol. Proc* 2006. 20 p. .

[Obot and Onyeukwu ()] 'Trend of rainfall in Abeokuta'. N I Obot , N O Onyeukwu . *J. Env. Iss. Agric. Dev. Count* 2010. 2006-2007. 2 (1) p. . (A 2-year experience)