Time Series Modelling of Rainfall in New Juaben Municipality of the Eastern Region of Ghana

Enock Mintah Ampaw
Department of Mathematics and Statistics
SAST, P. O. Box KF 981
Koforidua Polytechnic - Ghana.

Buckman Akuffo
Department of Mathematics and Statistics
SAST, P. O. Box KF 981
Koforidua Polytechnic - Ghana.
Email of corresponding author: **akuffobuckman@yahoo.com**

Seth Opoku Larbi
Department of Planning and Quality Assurance
Koforidua Polytechnic
Ghana.

Samuel Lartey
Department of Mathematics
Akro Senior High School
Krobo Odumase, Eastern Region - Ghana.

Abstract

The survival of virtually every organism is heavily dependent on water, one source of which is rainfall. Due to the agriculture base of the Ghanaian economy, rainfall forecast has become paramount in the wake of changing climatic conditions due to human activities. This paper therefore seeks to develop a time series model for rainfall pattern and its prediction for the New Juaben Municipality of the Eastern region of Ghana. Rainfall monthly data on New Juaben Municipality was used by applying the Box-Jenkins approach to develop the SARIMA (0, 0, 1) (1, 1, 1)$_{12}$ model, to forecast the rainfall statistics of the New Juaben Municipality for a year. From the results of the study, it was revealed that rainfall pattern for the next year will not be substantially different from the previous years despite some human and industrial activities that were believed to be affecting the pattern. It was therefore recommended that measures put in place to curb activities that seem to be detrimental to climate be continued.

Key Words: Rainfall, Seasonality, Differencing, Autoregressive and SARIMA.

1. Introduction

The lifeblood of every living organism is water, since no one can do without it. Individuals, societies, institutions and nations at large have managed their water resources judiciously to achieve specific objectives. Rainfall plays a critical role in the various aspects of our physical, chemical, biological processes and also serves as a regulator of the global environment (Anderson, 2000). The scientific study of climatology has revealed changes in the normal atmospheric conditions and weather patterns in the world in recent times.

In Ghana, climate change has affected the rainfall pattern greatly. Since 1995 the amount of rainfall recorded over the years has reduced drastically and its associated effects of rising temperature, drought, desertification, extinction of some plants and animal species, low crop yield and inconsistency in the pattern of the rainfall. Ghana has about 70% of the citizenry population engaged either directly or indirectly in rain dependent-agriculture (Kwamena, et al, 1995) hence there is the need to find ways of improving upon the climatic conditions of the country by identifying the causes of these changes so as to rectify them and ensure the welfare of the greater lot.

The forest resources of Ghana which was estimated to be 8.2million hectares in 1900 have been reduced to only 1.2 million hectares as a result of climate change and other activities such as bad mining practices, illegal logging, industrialization, urbanization as well as environmental pollution (Fredua, 1998). A look at the Eastern region of Ghana weather and rainfall patterns revealed that the region falls within the equatorial forest zone. It experiences the double maxima rainfall pattern namely the major and minor rainy seasons. The major rainy season starts from April and ends in July. On the other hand, the minor rainy season starts from September to October. Annual average rainfall is between 1,580mm and 1,780mm. Rainfall intensity, however, decreases towards the Volta basin. Mean monthly temperature ranges from as high as 30^{o}C in the dry season but declines to about 26^{o}C in the wet season (Kwamena, et al, 1995)

The New Juaben Municipality lies within the semi-deciduous (moderate amount of forest cover) forest zone in the Eastern Region of Ghana. The vegetation is dense in terms of tree coverage with most trees shedding their leaves in the dry season. It is estimated that the agriculture sector forms about forty percent (40%) of Ghana's economy. Again, about ninety (90%) of the workforce in the agricultural sector use mechanical methods in their operations. The agrarian economy in the new Juaben Municipality is characterized by over reliance, and dependence on rainfall, with few pockets of irrigation activities. The irrigation activities are done by some few farmers due to its accompanying financial implication, and thus their overall contribution tends to be insignificant. Hence, rain-fed farming tends to dominate in low income countries such as Ghana.

Again, market activities and construction are tremendously affected by rainfall statistics and distribution in a given geographical area. Thus, a model that is intended to model the expected rainfall statistics and distribution in the New Juaben Municipality will inure to the benefit of all rainfall users in and around the Municipality, and thereby contributing to the socio-economic development in the municipality and the nation at large. The prediction of rainfall by farmers and policy makers has been on by observation of months. However, human activities such as deforestation, mechanical mining, popularly known as "galamsy" and the indiscriminate dumping of refuse, improper management of both solid and liquid waste and the resultant effect of the depletion of the ozone layer, have all contributed to the changing pattern of rainfall in the municipality. The problem at stake is therefore how to accurately predict rainfall pattern in the New Juaben Municipality for farmers, and other policy makers for their activities.

The objectives of this paper are:

(i) To predict average expected monthly and yearly rainfall statistics in the municipality.

(ii) To determine if there is a linear trend of annual rainfall over time in the municipality for the period 1993 – 2011.

(iii) To determine if there is a change in the rainfall pattern in the municipality over the 18 year period.

(iv) To develop an ARIMA model that will be used to explain the time dependent structure of rainfall patterns.

In recent years, many technologies have been employed in predicting rainfall. The radar range equation relates the power-return from the target to the radar characteristics and parameters of the target. Greater details on weather radar theories can be found in different references such as: Skolnik (1970) for engineering and equipment aspects; Sauvageot (1982) Battan (1981), and Collier (1989) for meteorological phenomena and applications; Atlas (1964, and 1990) for general review; Rinehart (1991) for modem techniques; and Doviak and Zrnic (1993) for Doppler radar principles and applications. The radar measurements of precipitation can be summarized in two main steps: (1) conversion of the power return *(Pr) to* reflectivity factor (2) through the radar equation, and (2) conversion of the reflectivity factor into rainfall rate.

The estimation of rainfall from multi-sensors arose from the fact that rainfall measurements can be obtained from several sources. The most important sources are multiple radars, rain-gauge networks, and satellites. In addition, output from cloud models can be used as a rainfall estimator. Many studies have considered the problem of merging rainfall estimates from multi-sensors (Ahnert et al., 1983; Krajeweski, 1987; Seo et al., 19901, 2, and others). A statistical merging procedure (Krajeweski, 1987, and Seo et al., 1990) was employed in the NEXRAD (Next Generation Weather Radar) system in the United States.

This procedure merges data from multiple radars with rain-gauges by using the co-kriging optimal interpolation to produce a second stage (Stage-II) of rainfall accumulations products. Rainfall forecasting is usually classified into short, medium, and long range forecasting. In short range forecasting the lead time is less than 2 days ahead. Medium range forecasting is 2-14 days ahead, and long range forecasting is month's ahead (Bengtsson, 1985; and Collier, 1989). Short range forecasting is sub-classified into now casting, mesoscale forecasting, and synoptic scale forecasting (Browning, 1980). The development of statistical techniques arose as the difficulties of the physical modeling, owing to the diversity of climatic processes in different scales and the lack of detailed understanding of the rainfall producing mechanism, were recognized. These techniques are based on the so-called "black box regression" technique (Brier, 1950). These techniques try to relate the probability of precipitation (PoP) to the levels of some "explanatory variables".

Linear Statistical Models such as Autocorrelation functions, Spectral Analysis, Analysis of cross correlations; Linear Regression and Autoregressive Integrated Moving Average (ARIMA) have been studied for the applicability to flood forecasting. Solomatine et al., (2000) have found in their study that the use of stationary (ARMA) as well as non-stationary (ARIMA) versions of linear prediction techniques does not provide accurate predictions. Application of other linear stochastic methods has also resulted in inaccurate predictions, clearly indicating that linear statistical models do not accurately represent historical data and hence are not acceptable methods for a non-linear application such as flood forecasting.

In the auto-regressive processes where, persistence is present, that is the even outcome of the future is dependent on the present period magnitude. The Auto Regressive Moving Average (ARMA) processes represent a system of elements moving from one state to another over time. The MARKOV chain modeling approach has frequently been used for the synthetic generation of the rain-fall data. Thomas and Fiering (2001) used a first order MARKOV chain model to generate stream flow data. Srikanthan and Mohan (2000) used and recommended a first order MARKOV chain model to generate annual rain-fall data. Shamshad *et al* (1999) compared performance of stochastic approaches for forecasting river water quality. However a few studies have been done on the synthetic generation of rain-fall flow data using ARMA approach. ARMA approach is generally used for modeling and simulation of rain-fall flow data.

Min et al. (2010), used Autoregressive Integrated Moving Average (ARIMA) with integration model (also known as integration analysis) to evaluate the impact of different local, regional and global incidents of a man-made, natural and health character, in Taiwan over the last decade. The incidents used in this study were the Asian financial crisis starting in mid 1997, the September 21st earthquake in 1999, the September 11th terrorist attacks in 2001, and the outbreak of Severe Acute Respiratory Syndrome (SARS) in 2003. Empirical results revealed that the SARS illness had a significant impact, whereas the Asian economic crisis, the September 21st earthquake and the September 11th terrorist attacks showed no significant effect on air movements. Akuffo and Ampaw, (2013) used ARIMA in modelling Ghana's inflation from 1985 to 2011. Their model passed the relevant diagnostics checks and was used to forecast inflation for the year 2012. Their model was very accurate with predictive power of less than four (4) percent.

2. Material and Methods

2.1. Time Series Models

In time series model, the past behavior of a time series is examined to infer something about its future behavior instead of searching for effect of one or more variables on the forecast variable.

2.1.1. Autoregressive Average Model

Autoregressive average model represents current value of time series as combination of one or more previous values of the series. It shows the dependency of one value with its nearest previous values. Autoregressive process is a difference equation determined by random variables (difference equation shows current value of series as function of its previous values). Autoregressive model has order term, p that determines how many previous values are to be included in the difference equation to estimate current value.

The autoregressive AR (1), $p = 1$, includes only one previous value. It is a standard linear difference equation and written as:

$$Y_t = \phi_p Y_{t-p} + z_t \;\; for \;\; p = 0, \pm 1, \pm 2, \pm 3 \; \ldots \quad (1)$$

Where, u_t is error term and ϕ_p is parameter to be estimated. The p^{th} order AR time series, denoted by AR (p) is given by:

$$Y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \ldots \phi_p y_{p-1} + z_t \;\; for \;\; t = 0, \pm 1, \pm 2, \pm 3 \; \ldots \quad (2)$$

Where, $\phi_o, \phi_p \neq$ 0, and z_t are uncorrelated random variables.

Using difference equation, value of Y_t can be obtained from Y_{t-1}, value of Y_{t-1} is obtained from Y_{t-2} and so on. The **AR (1)** model is fitted with collected data by first estimating value of ρ. To estimate value of ρ least squares estimation method is used. It minimizes the sum of square of errors for the observed values with respect to ρ:

$$\frac{\partial}{\partial x} \sum_{t=2}^{2} (Y_t - \phi Y_{t-1})^2 = 2 \sum_{t=2}^{n} (Y_t - \phi Y_{t-1})(-Y_{t-1}) \ldots \quad (3)$$

Equating equation 3 to zero and solving it further gives the value of least square estimator for ϕ:

$$\hat{\phi} = \frac{\sum_{t=2}^{n} Y_t Y_{t-1}}{\sum_{t=2}^{n} Y_{t-1}^2} \ldots \quad (3a)$$

From estimated value of ϕ, distribution of error terms can be found as

$$\bar{z} = Y_t - \phi Y_{t-1} \ldots \quad (4)$$

Now using estimated value of ϕ and distribution of error data, the model can be fitted using equation (2).

2.1.2. Moving Average Process of Order q, MA (q)

A time series is influenced by random shocks in noisy environment. As a result current value of series is affected by the random shocks appeared in previous values. Moving average terms are used to capture the influence of previous random shocks in the future value.

First order moving average or *MA (1)* is a simple time series, given by

$$Y_t = z_t + \alpha z_{t=1} \ldots \quad (5)$$

This equation says, apart from mean, z, Y_1 is a weighted average of z_2 and z_1, Y_2, is a weighted average of z_2 and z_1 etc. The value of Y_t is defined in terms of random shocks z_t.

A Moving average of order *q*, MA (*q*) process X_t, is given by

$$Y_t = z_t + \theta_1 z_{t-1} + \ldots \theta_q z_{t-q} \ldots \quad (6)$$

Above equation (6) representing *MA (q)* process is always stationary**.** In fact *MA* process is inverse of *AR* model. The *MA* model is invertible if an *MA* model can be expressed as autoregressive (infinite order) model.

2.1.3. Autoregressive Moving Average Process, ARMA (p, q)

Autoregressive Moving Average model is formed by combining terms of AR and MA models. Autoregressive model or Moving Average can be used to approximate any stationary process with any degree of accuracy as desired. Combining equation (2) and (6), ARMA model of order *p* and *q* is formed,

$$Y_y = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + z_t + \theta_1 z_{t-1} + \ldots + \theta_q z_{t-q} \ldots \quad (7)$$

2.1.4. Autoregressive Integrated Moving Average Process - ARIMA (p, d, q)

The ARMA model assumes that the time series data is stationary (that is statistical properties of data do not change over time). But real data are often not stationary in nature. Time series data is made stationary by differencing process. The first order differencing process of time series X_t is defined as $X_t' = X_t - X_{t-1}$. ARMA time series which is made stationary by differencing process is known as Integrated Autoregressive Moving Average (ARIMA) model. ARIMA model is represented by three parameters: *p* order of autoregressive model, *d* order of differencing, and *q* order of moving average model. ARIMA model takes historical data and decomposes that data into an autoregressive (AR) process which maintains memory of past events, an Integrated (I) process which makes data stationary for easy forecast and a Moving Average (MA) process of forecast errors. It does not suffer from existence of serial correlation between the error residuals and their own lagged values. An ARIMA (*p, d, q*) model can be checked if it is a good statistical fit for data or not, using Akaike Information Criterion (AIC) and Schwarz Criterion (SC) method. Autocorrelation (AC) and partial autocorrelation (PAC) statistics help to determine the right parameters for ARIMA model.

2.1.5. ARIMA (p, d, q) with Exogenous Variable

An ARMA model simply adds in the covariate on the right hand side:

$$y_t = \beta x_t + \phi_1 Y_{t-1} + \phi_p Y_{t-p} - \phi_1 z_{t-1} - \ldots - \phi_q z_{t-q} - z_t \quad \ldots \quad \ldots \quad \ldots \quad \ldots. \quad \ldots \quad \mathbf{(8)}$$

Where x_t is a covariate at time *t* and β is its coefficient. While this looks straight-forward, one disadvantage is that the covariate coefficient is hard to interpret. The value of β is *not* the effect on y_t, when the x_t is increased by one (as it is in regression). The presence of lagged values of the response variable on the right hand side of the equation mean that β can only be interpreted conditional on the value of previous values of the response variable, which is hardly intuitive.

If we write the model in (7) using backshift operators, the ARMAX model is given by

$$\phi(B) y_t = \beta x_t + \theta(B) z_t \text{ or } y_t = \frac{\beta}{\phi(B)} x_t + \frac{\theta(B)}{\phi(B)} z_t \ldots \ldots. \quad \ldots. \quad \ldots \quad \ldots. \quad \ldots \quad \mathbf{(9)}$$

Where, $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$, and $\theta(B) = 1 - \theta_1 B - \ldots - \theta_p B^p$

2.2. The Box – Jenkins Methodology

The Box – Jenkins methodology seeks to transform any time series data to be stationary; and then apply the stationary process for forecasting by using past univariate time series process for future forecast with a host of selection and diagnostics tools. The process involves some three basic steps as discuss below.

2.2.1. Model Identification

This stage involves the specification of the correct order of ARIMA model by determining the appropriate order of the AR, MA and the integrated parts or the differencing order. The major tools in the identification process are the (sample) autocorrelation function and partial autocorrelation function (Akuffo and Ampaw, March 2013). The identification approach is basically designed for both stationary and non-stationary processes.

According to Table 1, spikes represent the line at various lags in the plot with length equal to magnitude of autocorrelations and these spikes distinguish the identification of a stationary and non-stationary process.
For stationary series if the series is MA (q) process, then this can be identified by an autocorrelation function (ACF) with zero at lags greater than *q*; and partial autocorrelations (PACF) tail off in exponential fashion. However, for an AR (*p*) process the partial autocorrelation (PACF) is zero at lags greater than *p*, autocorrelations (ACF) tail off in exponential fashion. Also in an ARMA processes, both the ACF and PACF will have large values up to *q* and *p* respectively, which tail off in an exponential fashion.

The complete frame work for the identification is as shown in the Table 1.

Table 1: Model identification criteria

Model	ACF	PACF
$AR(1)$	decays exponentially	single spike
$MR(1)$	single spike	decays exponentially
$AR(p)$	decays exponentially with damped oscillations	p spikes
$MR(q)$	q spikes	decays exponentially with damped oscillations
$ARMA(p, q)$	decays exponentially and damped oscillations	both decays exponentially

2.2.2. Model Estimation

Depending on the ACF and PACF of the sequence plots a model is run with appropriate software. The best fitting model must also have few parameters as much as possible alongside best statistics of the model according to the information selection criteria.

2.2.3. Model Checking

Model checking in time series can be done by looking at the residuals. Traditionally the residuals given by

$residual = observed\ values - fitted\ values.$

These residuals should be normally distributed with zero mean, uncorrelated, and should have minimum variance or dispersion, if indeed a model fits the data well. That is model validation usually consist of plotting residuals over time to verify the validation. A comprehensive procedure includes:

(i) Plot the residuals against time and inspect increasing (decreasing) variations which may suggest the need for data transformation.

(ii) Plot ACF and PACF of residuals

(iii) Plot residuals against fitted values and check for variations and correlation

(iv) Check the various t – ratio parameter estimates if any term(s) need to be dropped from the model.

(v) Check the correlograms derived from the residuals to determine whether additional terms are required.

Residual analysis can also be done through formal test using the Portmanteau test and other statistical tests.

3. Results and Discussion

3.1. Preliminary Analysis

The entire data span of 1993 - 2011, was divided into two; from January 1993 to December 2002, and January 2003 to December 2011. This was intended to reduce the tendency of modeling with structural breaks in the data since structural breaks was not formally tested. The summary statistics and time plots of monthly rainfall data was examined to check for the stability of the data. Table 2 shows the summary statistics of monthly rainfall data of the Municipality.

Table 2: Summary statistics of Municipality's rainfall from 1993 – 2011

Period	Statistic			Mean	Standard Deviation
	N	**Minimum**	**Maximum**		
Jan. 1993 – Dec 2002	120	0.20	380.20	121.10	84.22
Jan. 2003 – Dec 2011	108	0.70	399.70	120.80	83.88
Entire Period	**228**	**0.20**	**399.70**	**120.44**	**83.87**

From Table 2, there was a minimum rainfall of 0.20 mm, and maximum rainfall of 399.70 mm recorded in the New Juaben Municipality. A mean rainfall and a standard deviation of 120.44mm, and 83.88mm, respectively were recorded in the Municipality for the same period. It can also be observed from Table 2 that a minimum rainfall recorded from the New Juaben Municipality from the period January, 1993 to December, 2002 was 0.20 mm with a maximum of 380.20mm and a mean of 84.22. These vividly show an uneven distribution of rainfall amongst the months for the period January, 1993 – December, 2002. Again, the minimum and maximum rainfall figures of 0.70mm and 399.70 mm respectively, were recorded in the Municipality within the period January, 2003 and December, 2011. Furthermore, a mean rainfall of 120.8 mm, and a standard deviation of 83.88419 mm were realized for the period under consideration. This shows relatively slightly improvement over the former period per Table 2.

The value of the standard deviation of 83.88 mm shows that there was great dispersion of rainfall pattern amongst the months and years under study. That is rainfall was relatively high in some months, and in some years, and relatively low in some months showing a wide dispersion that may lead to non stationarity at level of rainfall figures in the Municipality. The time plot of the Municipality's rainfall pattern/distribution as depicted in Figure 1 revealed one which is irregular in nature, from the period 1993-2011. There were few isolated cases of tremendous rainfall in the Municipality. These isolated cases, as depicted by Figure 1, were experienced in 1993, 2001, and 2007, in these years, rains in the Municipality were relatively phenomenal.

The time plots with the summary statistics shows an indication that the data is not stationary at levels, exhibiting some seasonal behaviour and this must be formally tested.

3.2. The Structural Form of the Municipality's Rainfall Data

The examination of the summary statistics and visual inspection of the time plots gave evidence to the fact that the rainfall data may be non-stationary at levels and this formally check with the KPSS test for stationarity was given as:

$$H_o : stationary\ series$$
$$H_1 : non-stationary\ series$$

The results of the KPSS test are as presented in Table 3.

Table 3: Unit Root test for stationarity

KPSS Test Statistics for constant term	**0.882**
KPSS Test Statistics for constant term and a drift	**0.537**
1% Critical Value	0.216
5% Critical Value	0.146
10% Critical Value	0.119

From Table 3 it can be concluded that the rainfall data is non-stationary at 10%, 5% and 1% significant levels. The test statistic (0.882 for constant tern and 0.537 for constant term with a drift) are greater than the critical values at 10%, 5% and 1% significant levels as 0.119, 0.46 and 0.216 respectively.

It evident from Table 3, that, the null hypothesis is rejected. Hence, the rainfall data was non-stationary for both models. That is, a model with a drift, and one with a constants term and drift at critical value 1%, 5%, and 10%. The Box-Jenkins technique recommends transformation of the data by differencing.

3.3. Data Transformation and Stationary Check

The monthly data is differenced in attempt in achieving stationary. The differenced data is then tested for stationary with KPSS test. The result is presented in Table 4.

Table 4: Unit Root test for Stationary after differencing

	Ordinary Differencing	Seasonal Differencing
KPSS Test Statistics for constant term	0.523	0.135
KPSS Test Statistics for constant term and a drift	0.627	0.135
1% Critical Value	**0.216**	
5% Critical Value	**0.146**	
10% Critical Value	**0.191**	

The test statistics (**0.135** for constant tern and **0.043** for constant term with a drift) are less than the critical values of at 10%, 5% and 1% significant levels as **0.216, 0.146** and **0.191** respectively hence we fail to reject the null hypothesis and we conclude that the differenced data is stationary.

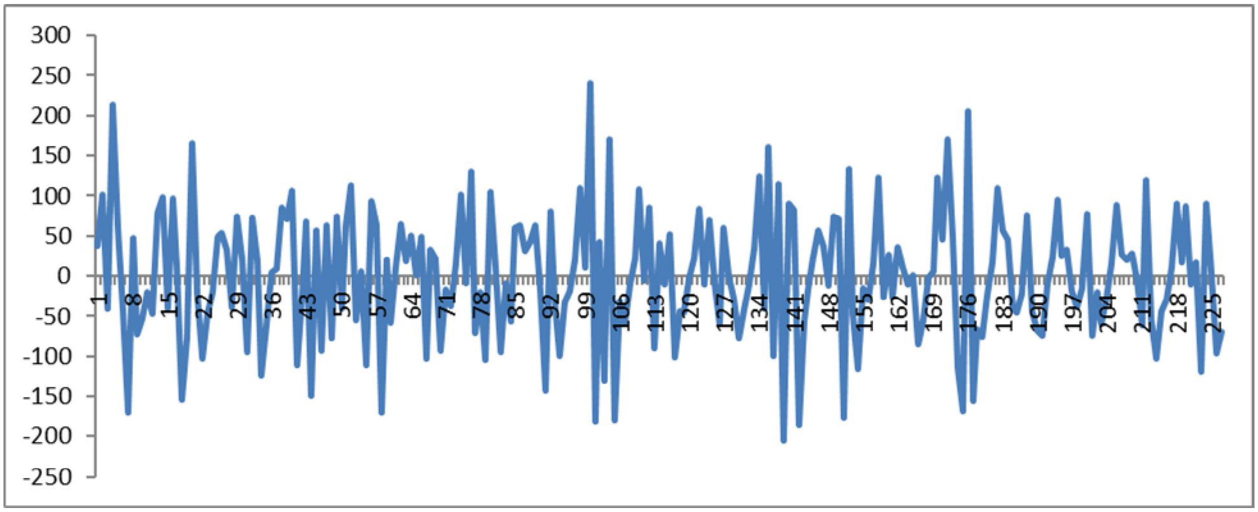

***Figure 2**: Time plot of the differenced monthly rainfall data in the Municipality (1993-2011)*

3.4. Model Identification

An assessment of the ACF and PACF as shown in Figure 2a and 2b indicate that there is seasonality as expected.

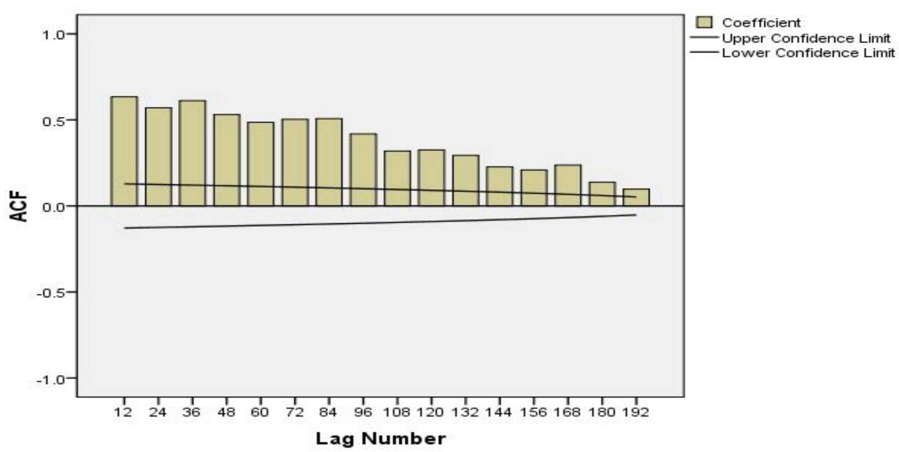

***Fig. 2a:** ACF time plot*

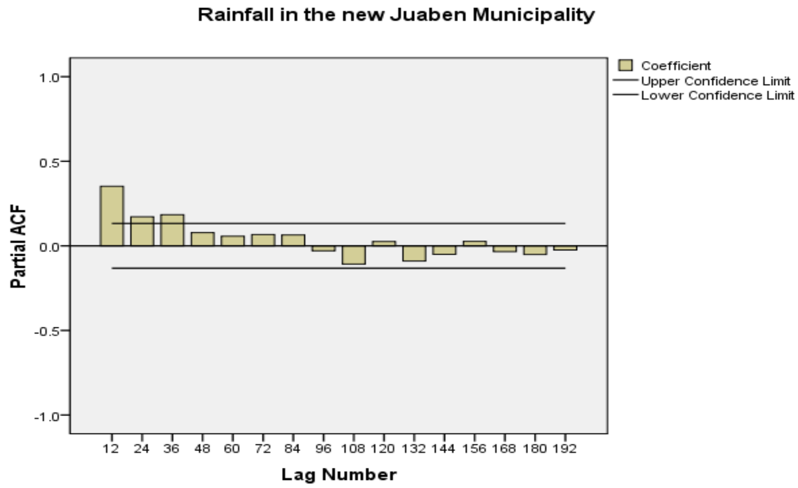

Figure 2b: *PACF plot*

The presence of exponential decay and damped oscillation were noticed in the figure, and this was an evidence of both AR, and MA parameters in the optimal model. Thus, ARMA processes with both ordinary and seasonal terms were considered. The spikes at lag 12, 24, and 36, as well as the exponential decay were indicative that, the model contained both AR and MA terms. Again, a large lag at 12, and a smaller one at 24, gave an indication of MA terms. Both the sample ACF, and PACF meant that a number of models be developed and the best one chosen for prediction, using some selection criteria such as the Akaike Information Criteria, and Bayesian Information Criteria.

3.5. Model Estimation

After a series of model tests, a family of models was produced from the R statistical package. The result of the models is as presented in Table 5.

Table 5: Estimated model of rainfall in the Municipality: SARIMA (0, 0, 1) (1, 1, 1)

Parameter	**Estimate**	**Standard Error**	**t-value**	**P (> |t|)**	**BIC**	**Durbin-Watson statistics**
SAR1: $\boldsymbol{\alpha_1}$	0.612	0.023	0.023	<0.0001	8.077	2.07
SAR2: $\boldsymbol{\beta_1}$	-0.453	0.005	0.541	<0.0001		
SMA1: $\boldsymbol{\beta_2}$	-0.635	0.095	0.376	<0.0001		

From Table 5, it can be seen that the best model is the **SARIMA (0, 0, 1) (1, 1, 1)$_{12}$** with normalized **BIC** of **8.077** and a Durbin Watson of 2.07.

The estimation of the parameters for the selected model is presented in Table 5. It can be seen from Table 5 that the model parameters are all statistically significant with a Durbin Watson statistic of 2.07 showing that the model has no autocorrelation problems.

The seasonal ARIMA with $p = d = q = 0,\ P = 2, D = 1 \ and \ Q = 1$(*two seasonal autoregressive, one seasonal difference and one seasonal moving average component*) is given as:

$$\left(1 - \alpha_1 \mathbf{B}^{12} - \alpha_2 \mathbf{B}^{12}\right) Y_t = \left(1 + \beta_1 \mathbf{B}^{12}\right) \varepsilon_t$$

Substituting, the parameters with a seasonal difference into, the model is derived as:

$$\Delta_{12} y_t - 0.612 \Delta_{12} y_{t-12} - 0.453 \Delta_{12} y_{t-12} = \varepsilon_t - 0.635 \varepsilon_{t-12}$$

$$y_t - y_{t-12} - 0.612\left(y_{t-12} - y_{t-24}\right) - 0.453\left(y_{t-12} - y_{t-24}\right) = \varepsilon_t - 0.635 \varepsilon_{t-12}$$

$$y_t - y_{t-12} - 0.612\,y_{t-12} + 0.612\,y_{t-24} - 0.453\,y_{t-12} + 0.453\,y_{t-24} = \varepsilon_t - 0.635\,\varepsilon_{t-12}$$

$$y_t - 2.065\,y_{t-12} + 1.065\,y_{t-24} = \varepsilon_t - 0.635\,\varepsilon_{t-12}$$

Hence

$$\hat{y}_t = 2.065\,y_{t-12} - 1.065\,y_{t-24} + \varepsilon_t - 0.635\,\varepsilon_{t-12}$$

This means the current rainfall can be predicted by a factor 0f 2.065 of the last twelve month's figure, less the last twenty-four month's figure with some innovation terms of the current year and the last twelve month's innovation term.

3.6. Diagnostics Checks

To check whether the model was correctly specified, the residuals were examined to be white noise. Table 6 shows that the model has a near zero mean and minimum deviation.

Table 6: Estimated model of rainfall in the Municipality

			Box-Ljung Statistic		
Lag	***Autocorrelation***	***Std. Error[a]***	***Value***	***df***	***Sig.[b]***
12	.634	.064	399.469	12	.000
24	.570	.062	736.533	24	.000
36	.612	.061	1110	36	.000
48	.531	.059	1440	48	.000
60	.487	.057	1730	60	.000
72	.504	.055	2041	72	.000
84	.507	.052	2344	84	.000
96	.419	.050	2616	96	.000
108	.319	.048	2836	108	.000
120	.325	.045	3040	120	.000
132	.294	.043	3221	132	.000
144	.227	.040	3374	144	.000
156	.209	.037	3503	156	.000
168	.238	.034	3652	168	.000
180	.138	.030	3764	180	.000
192	.097	.026	3833	192	.000

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Again, the plot of the autocorrelation function (ACF) of the residual must die out after lag one (1) as shown in Figure 3. Also, the ACF residuals must lie within control limit. All these affirm the fact that the selected model is correctly specified and does not violation any serious assumption of time series modelling as specified by the approach adopted.

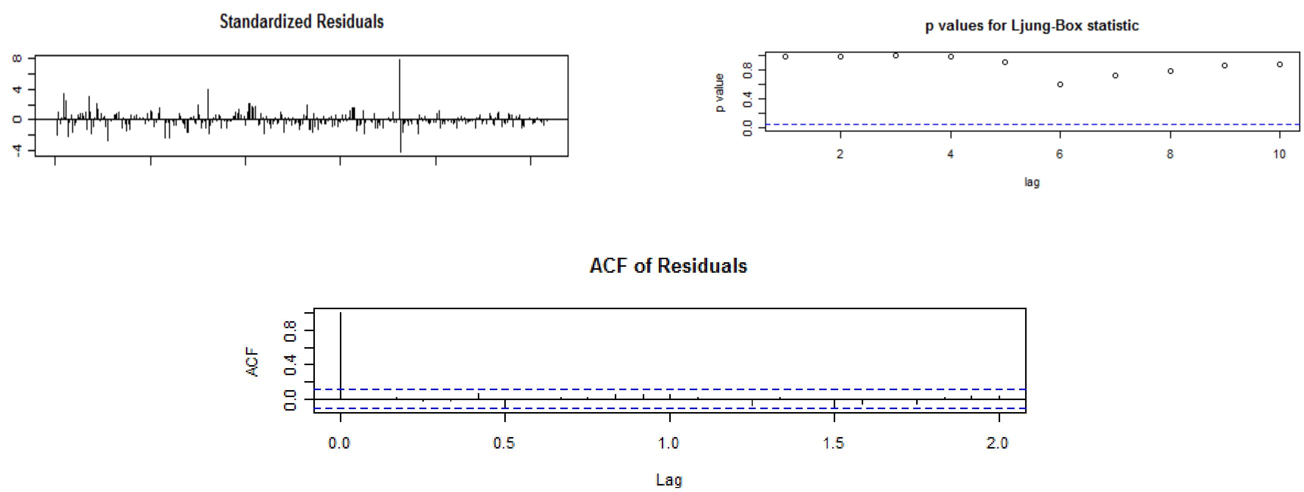

Fig. 3: *Plots of residuals*

3.7. Forecast

The selected model was used to forecast rainfall in the Municipality for twelve month for the year 2012. The result of the forecast is as shown in Table 7.

Table 7: One year forecast of rainfall distribution in the Municipality from SARIMA (0, 0, 1) (1, 1, 1)$_{12}$ model.

Date	Forecast (MM)	Actual (MM)	Difference	Lower 95% CL	Upper 95% CL
January 2012	**4.65**	**3.85**	*0.80*	*-83.60*	*92.90*
February 2012	**33.14**	**42.05**	*8.91*	*-56.86*	*123.13*
March 2012	**167.32**	**163.51**	*-3.81*	*27.61*	*211.03*
April 2012	**186.13**	**181.01**	*-5.12*	*39.74*	*226.52*
May 2012	**219.16**	**225.04**	*5.88*	*124.12*	*314.20*
June 2012	**200.73**	**213.64**	*12.91*	*104.07*	*297.39*
July 2012	**194.10**	**198.67**	*4.57*	*95.84*	*292.36*
August 2012	**111.26**	**108.98**	*-2.28*	*11.43*	*211.09*
September 2012	**187.65**	**180.73**	*-6.92*	*86.18*	*266.94*
October 2012	**158.55**	**172.30**	*13.75*	*55.65*	*261.45*
November2012	**67.76**	**71.45**	*3.69*	*-36.73*	*172.08*
December 2012	**20.68**	**18.65**	*-2.03*	*-75.20*	*136.50*

Table 7 shows the forecast and actual rainfall data, together with 95% confidence interval, generated from the estimated rainfall model, **SARIMA (0, 0, 0) (2, 1, 1)$_{12}$**, of the municipality. From Table 7, the forecast values lie within the lower and upper limits. The difference of the actual and forecasted figures also show very marginal differences. The fact that the differences have both negatives and positives figures shows that the model is neither over forecasting nor under forecasting. The forecast and actual plot is as shown in Figure 4.

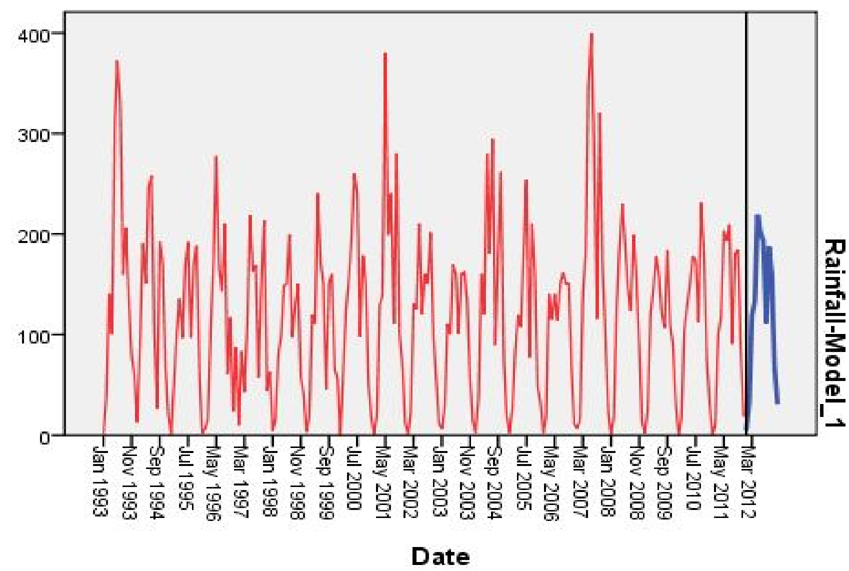

Figure 4: *Actual and Forecast of Rainfall in the Municipality (1993 – 2012)*

The forecast shows a steady trace of the actual values over time. This also means that all things being equal, rainfall pattern in the New Juaben Municipality is not going to witness any substantial change in 2012 as recorded earlier.

4. Conclusion

The paper sought to develop a time series model based on recorded rainfall data from 1993-2011, of the New Juaben Municipality for predicting the rainfall distribution and pattern of the New Juaben Municipality. A preliminary check on the data shows that the data is non-stationary, but stationarity was achieved after first seasonal differencing. The best model was identify as **SARIMA (0, 0, 0) (2, 1, 1)$_{12}$** with the smallest BIC of 8.077 and a Durbin-Watson statistics of 2.07.The best selected model was identified as

$$\hat{y}_t = 2.065\, y_{t-12} - 1.065\, y_{t-24} + \varepsilon_t - 0.635\,\varepsilon_{t-12}\,.$$

Thus, the current rainfall can be predicted by a factor of 2.065 of the last twelve month's figure, less the last twenty-four month's figure with some innovation terms of the current year and the last twelve month's innovation term.

References

Akaike H.,(1976), "*Canonical Correlation Analysis of Time Series and use of Information Criterion," in System Identification*: Advances and Case Studies, Academic Press, New York.

Akufffo B., and Ampaw E. M, (2013). *International Journal of Science and Technology Education, Mathematics Modelling and Theory*, Vol. 3 No. 3: 10 - 26

Anderson B.O., et al, (1979), *Optimal Filtering*, Prentice Hall, Engle Wood Cliffs, New York.

Ansley C. F., R. Kohn, (1985), *Filtering, and Smoothing in State Space Models with. Incomplete Specified Initial Condition*, Ann Statist.

Ansley C.F., P. Newbold, (1980), *Finite Sample Properties of Estimation for Autoregressive Moving-Average Models*, J Econometrics.

Astrom, K. J., T. Bohlin (1966), "Numerical Identification of Linear Dynamics System from Normal Operation Record," In the Theory of Self-Academic Control System, P. H. Plenum Press, New York.

Banco De Espana (1999), *Bolclin Economica Monthly Bulletin*, Madrid: Banco De Espana.

Bell W.R. and Hillmer, S.C.(1991*), Initializing the Kalman Filter for Non stationarity Time Series Models,* Journal of Time Series Analysis 12, 283-300.

Bell, W.R. (1981), *Signal Extraction for Non-stationary Time Series,* Annals of Statistics 12, 646-661.

Bernard G. A., G. M. Jenkins,(1962*), Likelihood Inference and Time Series,* J. Roy Statist.

Black, T.I., (1994), *The New NMC Mesoscale Eta Model: Description and Forecast Examples,* Weather and Forecasting, 9(2):265-278.

Box, G. E. P., G. M. Jenkins, (1969), "*Models for Forecasting Seasonal and Non seasonal Time Series*," in Spectral Analysis of Time Series, Wiley, New York.

Box, G.E.P., and Jenkins, G.M. (1970), *Time Series Analysis: Forecasting and Control,* San Francisco: Holden-Day.

Box, G.E.P., and Tia, O.G.C., '*Intervention Analysis with Applications to Economic and Environmental Problem.' Journal of the American Statistical Association* 70, 71-79.

Brier, G.W.,(1950), *Verification of Forecasts Expressed in Terms of Probability, Monthly Weather Review,* 78(1): 1-3.

Brockwel, L. P. and Davis, R. (1987), *Time Series: Theory and Methods,* Berlin: Springer Verlag.

Brown, R.G. (1962), *Smoothing, Forecasting and Prediction of Discrete Time Series,* Prentice Hall, Englewood Cliffs, New York.

Bruton, J.M., McClendon, R.W. and Hoogenboom, G. (2000), *Estimating Daily Pan Evaporating with Artificial Neural Networks, Transactions of ASAE,*43(2):491-496.

Burman, J. P. (1980), '*Seasonal Adjustment by Signal Extraction', Journal of the Royal Statistical Society*. A 143, 321-337.

Burridge, P. and W. Allis K. F. (1984), 'Unobserved Component Models for Seasonal Adjustment Filters', Journal of Business and Economic Statistics 2, 350-359.

Carrol K.L., and Maloney, J.C., (2004), Improvement in Extended-Range Temperature and Probability of Precipitation Guidance, Symposium on the 50th Anniversary of Operational Numerical Weather Prediction, Am. Meteor. Soc., College Park MD.

Chen, C. and Liu, L. M. (1993), '*Joint Estimation of Model Parameters and Outlier Effects in Time Series'*. Journal of the American Statistical Association 88, 284-287.

Congalton, R.G., (1999), *A Review of Assessing the Accuracy of Classification of Remotely Sensed Data*, Remote Sensing of Environment 37:35-46.

De Jong P. (1991), *The Diffuse Kalman Filter.'* Annals of Statistics 19, 1073-1083.

Eurost A. T. (1991), '*Monthly Panorama of European Business Luxembourg:* Official Publications of the European Communities.

Findley, D.F., Monsell, B. C., Bell, W. R., Otto, M.C., and Chen, B. C. (1998), '*New Capabilities and Methods of the X12 ARIMA Seasonal Adjustment Program'(With discussion)*. Journal of Business and Economic Statistics 16, 127-177.

Fisher R. A., (1956), *Statistical Methods, and Scientific Inference*, Oliver and Boyd, Edinburg. Gebhard K., Jurgen Wolters, (2007), *Introduction to Modern Time Series Analysis.* Springer Verlag, Berlin.

George E.P., Box. G. Jenkins, et al, (2004), *Time Series Analysis: Forecasting and Control*, Prentice-Hall International Inc. New Jersey.

Glahn, H.R., and Lowry, D.A., (1972), *The Use of Model Output Statistics in Objective Weather Forecasting*, Journal of Applied Meteorology, 11(8):1203-1211.

Hall, T., Brooks, H.E., and Doswell, C.A. (1999), *Precipitation Forecasting Using a Neural Network, Weather and Forecasting.* 14(3):338-343.

Hamilton D., (2000), *Time Series Analysis*, Princeton University Press, Princeton., USA.

Harvey A. C., (1989) *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge University Press, UK.

Harvey, A. C., and A Scott, (1994), *Seasonality in Dynamic Regression Models,* Economic Journal 104: 1324-1345.

Harvey, A.C., and P.H.J. Todd, (1983) *Forecasting Economic Time Series with Structural and Box-Jenkins Models*, Journal of Business and Economic Statistics.

Harvey, A.C., and S. Peters (1998*), Estimation Procedures for Structural Time Series Model,* LSE Econometrics Programs Discussion Paper.

Hatch, U., Paudel, K., Cruise, J., Lamb, M., Masters, M., Limaye, A., and Perkey, D (200). *Hydraulic-Economics Evaluation of Use and Value of Irrigation Water."* Department of Agricultural and Rural Sociology, Auburn University.

Hess, U., J. Syroka, (2005), *Weather-based Insurance in Southern Africa: The case of Malawi Agriculture and Rural Discussion Paper 13*. Washington, D. C.: World Bank.

Houck, J.P and M.E. Ryan. (1972)*"Supply Analysis of Corn in the United States: The Impact of Changing Government Programs."* Amer. J. Agar. Econ. : 184-91.

Houston, J.E., Christopher S. McIntosh, Paul A. Stavriotis, and Steve Turner. *Leading Indicators of Regional Cotton Acreage Response: Structural and Time Series Modeling Result,* Journal of Agricultural and Applied Economics, 31, 3 (1999):507-517.

Jari, D.F. and Afzet, M. M. (1999), *Analysis and Modeling of Seasonal Time Series, Seasonal Analysis of Economic Time Series,* Washington D.C: U.S. Dept. of Commerce, Bureau of the Census, 309-331.

Kuk, K.C., Ball, J.E, and Sharma, A (2000), *A study of Optimal Lag and Spatial Inputs for Artificial Neura.*

Kuligowski, R.J. and Barros, A.P., 1998, *Localized Precipitation Forecasts From a Numeral Weather Prediction Model Using Artificial Neural Network*, Weather Forecasting Journal. U.S.A.

Kwamena Dickson, George Benneh (1995). *A New Geography of Ghana*. Longman Group, England.

Makaudze, E. (2005). *Do Seasonal Climate Forecast and Crop Insurance Matter for Smallholder Farmers in Zimbabwe*? Department of Agriculture, Ohio State University, USA. *Network For Rainfall Forecasting*, Journal of Hydrology.

R.B Bunnet, R.B., Olatunde Okunrotifa (1994), *General Geography In Diagrams For West Africa,* Longman Group, England.

Robert F. Engle, Tim Bollerslev, (1986), *Modeling the Persistence of conditional Variances*, Econometric Reviews.

Schaefer, J.T (1990), *The Critical Success Index as an Indicator of Warning Skill, Weather and Forecasting.*

Shank, D.B., McClendon, R. W., Paz, J.O, and Hoogenboom G (2008), *Ensemble Artificial Neural Networks for Prediction of Dew Point Temperature, Applied Artificial Intelligence, 22(6)*:523-542.

Smith, B. A., MaClendon, R. W. and Hoogenboom, G., (2006), *Improving Air Temperature Prediction with Artificial Neural Networks*, International Journal of Computational Intelligence.

Specht, D.F., (1990), *Probabilistic Neural Networks*, Neural Networks, 3(1):109-118.