UNIVERSITY FOR DEVELOPMENT STUDIES

MARKOV REGIME-SWITCHING AUTOREGRESSIVE MODEL FOR MODELING RAINFALL IN THE UPPER EAST REGION OF GHANA

DANIEL JNR SOROGO

2021

UNIVERSITY FOR DEVELOPMENT STUDIES

MARKOV REGIME-SWITCHING AUTOREGRESSIVE MODEL FOR MODELING RAINFALL IN THE UPPER EAST REGION OF GHANA

BY

DANIEL JNR SOROGO (PGD STATISTICS)

(UDS/MAS/0005/19)

Thesis Submitted to the Department of Statistics, Faculty of Mathematical Sciences, University for Development Studies in Partial Fulfillment of the Requirements for the Award of Master of Philosophy Degree in Applied Statistics

SEPTEMBER, 2021

DECLARATION

Student

I hereby declare that this thesis is the result of my own original research and that no part of it has been presented for another certificate in this University or elsewhere.

Candidate's Signature:……..……..… Date:……..…

Name: Daniel Jnr Sorogo

Supervisors

We hereby declare that the preparation and presentation of the thesis was supervised in accordance with the guidelines on supervision of thesis laid down by the University for Development Studies.

Main Supervisor's Signature: ……..… Date: ……..……

Name: Prof. Albert Luguterah

Co-Supervisor's Signature: …… Date: ……..…….

Name: Mr. Abdul Ghaniyyu Abubakari

ABSTRACT

The economy of Ghana largely depends on agriculture for its growth. In Ghana, agricultural productivity largely relies on rainfall. This study seeks to explore the rainfall pattern using data from the Ghana Meteorological Agency from January 1999 to December 2019 for the Navrongo and Bolgatanga weather substations. A univariate two-state Markov switching autoregressive model was used to characterize the distinct seasonal regime behavior of the rainfall in the study area. The data sets were found to be non-stationary. Hence, were differenced to attain stationarity. MS (2)-AR (1) was obtained as the best model for the Navrongo data as it had the least of goodness-of-fit measures. Also for the Bolgatanga data sets, MS (2)-AR (1) was obtained as the best model. The transitional probabilities for the Navrongo weather substation are $P_{11} = 0.80$, $P_{12} = 0.20$, $P_{21} = 0.35$, and $P_{22} = 0.65$ with a 5.1 months expected duration of rainfall and at least 3.05 months of dry season every year. The transitional probabilities for the Bolgatanga weather substation are $P_{11} = 0.82$, $P_{12} = 0.18$, $P_{21} = 0.27$, and $P_{22} = 0.73$. The expected duration for rainfall in the study area is 5.5 months and that for low or no rainfall is estimated to be at least 3.7 months for every year. Based on the variations in the expected duration for the various regimes, it is recommended that government assists farmers with early maturing seedlings and drought resistant seedlings. Due to uneven distribution of the rainfall pattern observed, the study recommends farmers practice good water management in their farms as it is essential to crop survival and the maximization of crop yield potential.

ACKNOWLEDGEMENTS

I wish to express my profound gratitude to many who assisted, directed and encouraged me to finish this work. Many thanks to my supervisor, Professor Albert Luguterah and my co-supervisor, Mr. Abubakari Abdul Ghaniyyu for their professional guide, endless encouragement and patiently guiding me through this process, in addition to providing constructive feedback on endless drafts.

I sincerely appreciate the invaluable direction and encouragement of the Head of Department, Dr. Suleman Nasiru. He played an instrumental role to ensure this work was successfully completed. To all my lecturers, my colleagues and anyone that played a role in this my academic life, I say thank.

I would like to also sincerely thank my parents and my siblings for their prayers, encouragement and their financial support. Finally, I am indebted to my wife Mrs. Joyce Laadi Sorogo for her support, advice, and love to enable me go through the programme successfully.

DEDICATION

I dedicate this work to my lovely wife Mrs. Sorogo Jnr and my parents Mr. and Mrs. Daniel Sorogo.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

ACF	Autocorrelation Function
ADF	Augmented Dickey-Fuller
AIC	Akaike's Information Criterion
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
CC	Climate Change
GHGs	Greenhouse Gases
GoG	Government of Ghana
GSS	Ghana Statistical Service
HQC	Hannan and Quinn criterion
IPCC	Intergovernmental Panel on Climate Change
KPSS	Kwiatkowski Phillips Schmidt–Shin
MA	Moving Average
MLE	Maximum Likelihood Estimation
MS	Markov switching
MSAR	Markov Switching Autoregressive

PACF	Partial Autocorrelation Function
PP	Philip-Perron
SARIMA	Seasonal Autoregressive Integrated Moving Average
SD	Standard Deviation
UNFCCC	United Nation Framework Convention on Climate Change
VAR	Vector Autoregressive
WMO	World Meteorological Organization

CHAPTER ONE

INTRODUCTION

1.1 Background of Study

A number of challenges have plagued the world in recent times, with climate change being one of the biggest challenges of the 21st century. The impact of climate change is global in scope and unprecedented in scale, and as a global challenge, requires global effort to carve solutions to combat (UNFCCC, 2007). The threat of climate change is multidimensional and its impacts transcend continents and national boundaries. The Intergovernmental Panel on Climate Change (IPCC) projects that if emissions of greenhouse gases continue to rise at their current pace, the world will be faced with a catastrophic future. These catastrophes among others are shifting in weather patterns that threaten food production and food security and the rising sea levels that increase the risk of flooding. Others include shifts in crop and plants growing seasons, loss of biodiversity, as well as increased frequency and intensity of extreme weather events such as heat waves, storms, floods and droughts (Wuebbles et al., 2017). Developing countries, particularly those in Africa, and generally the poor and marginalized, will be the worst affected by the impact of climate change, albeit the largest share of historical and current global emissions of greenhouse gases which are the key driving forces behind climate change, emanates from the developed and industrialized countries (Awojobi and Tetteh, 2017). This can be as a result of the economies of developing countries heavily dependent on sectors that are sensitive to climate change such as agriculture, forestry, hydro-power

among others. Also, African countries are less likely to adapt easily to the increasing and devastating impact of climate change hence makes it the most venerable (Bruckner, 2012).

Agriculture, a sector sensitive to global climate change, is one among the main contributors to the expansion of the Ghanaian economy in terms of contribution to Gross Domestic Product (GDP), foreign exchange earnings, employment and supply of crucial inputs to agro based industries among others. Albeit the contribution of agriculture to the GDP of the Ghanaian economy over the years have declined, agriculture contributed about 20% to the GDP in the year 2019 (GSS, 2019). According to the Ghana Living Standard Survey Round 6 (GLSS 6), agriculture provided employment to approximately 49% of the country's population while the contribution of employment to the rural dwellers was over 70% of the rural population (GSS 2012). Agriculture has also ensured food security in the country, even though there are cases of food insecurity (ISSER, 2013).

According to ISSER (2013), an average of 9% of the nation's population experience food insecurity annually. This figure is worst at the regional level as the then three northern regions (Northern, Upper East and Upper West) recorded 17%, 20% and 13% cases of food insecurity respectively. A further analysis of these phenomena indicates that these regions record average rainfall below the national average (ISSER 2013). The total output from agriculture can therefore be said to depend on the total amount of rainfall received and the nature of the distribution of the rainfall. That is, the total amount and distribution of rainfall

during the growing season are critical to crop yields. Analysis of rainfall and crop yields in Ghana indicates that rainfall variability, uneven distribution, insufficient amount and delay in onset of rains contribute to decline in crop yields with reasonable amount in especially, the northern part of the country.

Rainfall variability has historically been found to be a major cause of food insecurity. This is because the agricultural sector is facing increased and continued risks of climate change (Owusu and Waylen 2013). Generally, agriculture in Ghana is rain fed, with over 80% of agriculture productivity in Ghana depending on rainfall (De Bon et al., 2010). Not only does rainfall variability affect the agriculture sector, but also affects the energy sector. The total grid electricity generation of Ghana is 14,069 Gigawatt-hours (GWh) of which about 40% of the source of this installed electricity capacity is comprised of hydro power (GoG, 2018). These hydro power sources include the Akosombo, Kpong and Bui. These plants rely on water for power generation and therefore gets their sources from rainfall. It therefore suggest that insufficient rains will hamper the generational abilities of these plants. In 2015, the Akosombo dam felt below its minimum operational capacity due to insufficient rains which hampered electricity generation (GoG, 2015).

Ghana is situated in one of the world's most complex climatic regions, affected by tropical storms, and the influence of the Atlantic Ocean and the Sahel (GoG, 2013). Ghana is not immune to these consequences of climate change. Analysis of Ghana weather data from 1960 to 2000 indicates a decline in the mean annual rainfall (GoG, 2013). The rainfall in Ghana decreases from the southern belt to

the northern belt. Over the years, Ghana has seen the average annual rainfall decrease with torrential rains and extreme events resulting in storm damage, flooding, and sea-level rise (GoG, 2013).

1.2 Statement of the Problem

From when the farmer tills the land, to sowing and harvesting, countless decisions that affects their crop yields are made on daily basis, weekly and even months. These decisions such as when to sow, when to apply fertilizer and when to spray among others are geared towards maximizing their yields and minimizing cost. One decision that is beyond the control of the farmer however is the rainfall. The importance of rainfall to the farmers therefore cannot be overemphasized. The total output from agriculture is said to depend directly on the total amount of rainfall received and the nature of the distribution of the rainfall (Owusu and Waylen 2013). That is, the total amount and distribution of rainfall during the growing season are critical to crop yields. Analysis of rainfall and crop yields in Ghana indicates that rainfall variability, uneven distribution, insufficient amount and delay in the onset of rains contribute to decline in crop yields with reasonable amount in especially, the northern part of the country. Over the years, Ghana has seen the average annual rainfall decrease with torrential rains and extreme events resulting in storm damage, flooding, and sea-level rise (GoG, 2013). The rainfall in Ghana especially decreases from the southern belt to the northern belt (GoG, 2013). Rainfall variability has historically been found to be a major cause of food insecurity and as such, the agricultural sector is facing increased and continued risks of climate change (Owusu and Waylen 2013). Data from Ghana statistical

service (2014) indicates that over 70% of the rural dwellers in the Upper East Region engage in various forms of agriculture productivity for survival. A significant rainfall variability will therefore have a negative impact on the crop yield and income of every household that rely on agriculture. Over 80% of agriculture productivity in Ghana rely on rainfall with little alternatives such as irrigational schemes (De Bon et al., 2010). Modeling and forecasting rainfall pattern will therefore be of immersed importance to the farmers in the Upper East Region of the country to reduce these risks. This study therefore seeks to bridge this gaps by employing a Markov regime-switching model for modelling and forecasting rainfall in the Upper East Region of Ghana.

1.3. Objectives of the Study

This section presents the general and the specific objectives of the study.

1.3.1 General Objectives

The general objective of the study is to model the rainfall distribution in the Upper East Region using Markov Regime-Switching Autoregressive Model.

1.3.2 Specific Objectives

The specific objectives of the study include:

i. To investigate the pattern of the rainfall in the Upper East Region.
ii. To develop a Markov regime-switching autoregressive model for each meteorological station in the Upper East Region.
iii. To estimate the expected duration of rainfall in the Upper East Region.

1.4 Significance of the Study

In the face of uncertainties surrounding climate change, it is imperative to control and reduce the emission of greenhouse gases which are the key drivers of climate change, while also finding means to cope with the impact of the phenomena. Ghana over the years has made tremendous progress by attaining a middle-income status as defined by the World Bank (GoG, 2013). This progress however is under serious threat from the impact of climate change, as Ghana is vulnerable to climate change and variability due to reliance on sectors that are sensitive to climate change, such as agriculture, forestry, and energy production. To be able to adapt, citizens need to find the best-fit plan to cope with the impact of climate change. These could include among others a model that could guide them in predicting the rainfall pattern and variability. The research findings would therefore be crucial to the indigenes of the Upper East Region who are predominantly farmers as it will enable them to understand the rainfall variability in the area and to be able to predict the rainfall. The research work is also aimed at contributing to existing knowledge in the field of academia.

1.5 Limitation of the Study

The following were the limitations of the study;

i. Due to time and resource constraint at the researcher's disposal, the study was limited to only two meteorological stations; Navrongo and Bolgatanga in the Upper East Region of Ghana.

i. The monthly rainfall data used for the study were secondary data and might contain some human errors.

ii. Lack of daily rainfall data at the disposal of the researcher made it difficult and impossible to study the monthly distribution of the rainfall.

1.6 Structure of the Thesis

This thesis is divided into five chapters. Chapter one contains the introduction and background of the study. Chapter two comprises of literature review. The methodology used in the study is contained in chapter three. Chapter four presents the analysis and discussion of results while chapter five contains the summary, conclusions and recommendations.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

There has not been scarcity of studies carried out into rainfall modeling and forecasting in Ghana, Africa and the world over. Various researchers have used different approaches to model the rainfall patterns and forecast the rainfall regimes. This chapter is dedicated to discuss the works of previous researchers.

2.2 Modeling and Forecasting Rainfall

The importance of rainfall cannot be overemphasized. In developing countries like Ghana, key sectors like the agriculture sector and the energy sector among others, contribute enormously to the Ghanaian economy rely heavily on rainfall. It is estimated that over 80% of crop production in Ghana is rain-fed (De Bon et al., 2010). Due to the impact of climate change on rainfall variability, being able to model and forecast the rainfall pattern will reduce the risk associated with the rainfall.

Asamoah et al. (2020) in analyzing the temporal description of annual temperature and rainfall in the Bawku area used data in the area from 1976 to 2015. The Mann–Kendall (MK) was used to test the trends in the rainfall pattern in the area and the standardized Precipitation Index (SPI) was also used to describe the anomalies of total annual rainfall received in the area. The study concluded that the Bawku area experience a single rainfall regime, with the onset of the rains in late March and ends in October.

Nyatuame et al. (2018) in their research to model the annual rainfall and maximum temperature over Tordzie watershed using the Stochastic autoregressive integrated moving average (ARIMA) model, used data from 1984–2014. The best fitted seasonal autoregressive integrated moving average (SARIMA) model for the rainfall in the area was $(1,1,1)\mathrm{x}(0,1,2)_{12}$ with a BIC of 9.051574. The study findings revealed that rainfall is generally declining slowly while temperature in rising. They concluded that the consequences of the decreasing rainfall are an increase in drought leading to food insecurity.

Filder et al. (2019) proposed the application of SARIMA models in analyzing and forecasting the monthly rainfall patterns. The study revealed that SARIMA $(1,1,1)\ (0,1,2)_{12}$ model with an AIC score of 9.99356 was deemed to be appropriate for the prediction of the rainfall pattern.

Ampadu et al. (2019) in their study examined the distribution of rainfall in the Upper East Region of Ghana from 1976 to 2016. The findings from the study revealed that the selected stations in the upper East Region have uni-modal rainfall distribution and that the rain mostly starts in May and ends in September. High precipitation occurs in July, August and September, with August recording the highest amount with a low variability, indicating the reliable occurrence of precipitation within this period of the year.

Nkrumah et al (2014) examined the rainfall variability over Ghana with model versus rain gauge observation scenario. The study made use of daily rainfall data from 1990 to 2008 from six synoptic weather stations. The study results indicated that rain gauges located at areas with unimodal rainfall distribution have rainfall

periods of April to October with September and October recording the highest amount of rainfall. This finding is consistent with Asamoah et al. (2020). The middle belt however which has a bimodal rainfall regime and also peaks in August, exhibits similarities in the nature of their rainfall patterns. The first peak for the bimodal stations in the southern belt was observed to be higher than the second peaks. Observation from the Regional Climate Model, Version 3 (RegCM3) data in both Accra and Axim revealed that there was an underestimation of rainfall amount at both stations except for Accra which recorded an overestimation of rainfall amount in September and October of about 3 mm and 8 mm respectively. The study also showed that similar RegCM3 data in Kumasi shows a similar rainfall pattern to that of the rain gauge data but with a high underestimation of the rainfall amount. Other stations such as Wenchi, recorded its first average maximum rainfall in June with 94 mm of rainfall and its second average maximum recorded in September with 119 mm of rainfall, which is still an underestimation in the amount of rainfall.

Iddrisu et al. (2016) in modeling the trend of flows with respect to rainfall variability employed the vector autoregression approach. The study results revealed that approximately 81% of the variability in the trend of rainfall has been accounted for by past innovations in rainfall figures at Bui while a significant amount of about 19% of the variability in the trend of rainfall have been explained by past innovations in flow values at Bui. Also, the study further revealed that approximately 78% of the variability in the trend flow has been explained by previous innovations flow at Bui, and on the other hand approximately 22% of

the variability in the trend of flow has been accounted for by past innovations in rainfall figures at Bui.

Issahaku et al. (2016) examined the rainfall and temperature changes and variability in the Upper East Region of Ghana using data from 1954 to 2014. Multiple linear regressions were the preferred model employed to model the rainfall and temperature variability. The decomposition procedures were also used to make forecasts of temperature and rainfall in the Region. In selecting the best model for the rainfall and temperature variability, the maximum absolute percentage error (MAPE), mean absolute deviation (MAD), and mean squared deviation (MSD) were approaches used. The study results showed that the Upper East Region has a unimodal rainy season usually from April to October. With respect to the rainfall variation, the study results revealed that there were variations in rainfall between the months. The variations in monthly rainfall and in both night and day temperatures determine the rainy and dry seasons in the Upper East Region. The rainfall regime indices indicated that the month of August is the peak of the rainy season with indices of 185.632 mm while January is the peak of the dry season with indices of -77.447 mm.

Adenomon et al. (2013) proposed the vector autoregressive (VAR) approach to modeling the dynamic relationship between rainfall and temperature time series data in Niger State, Nigeria using data from 1981 to 2010. Augmented Dickey-Fuller (ADF) was used to test for the stationarity of the series which revealed that the rainfall and temperature time series data were both stationary. This assertion was supported by VAR stability condition. Evidence from the study showed that

over 86% of the variance in Rainfall appears to have been explained by innovations in Rainfall and 8% was explained by innovations in temperature. It was further revealed that a great percentage (over 91%) of the variance in temperature appears to have been accounted for by innovations in temperature, while over 13% was explained by innovations in rainfall.

Chowdhury et al. (2017) proposed the development and evaluation of a stochastic daily rainfall model with respect to the long-term variability. The study aimed at developing a stochastic rainfall generation model that can represent not only the daily variability but also the longer variability such as inter-annual, decadal of observed rainfall. To achieve this, the study developed a Markov chain model with two-state Markov chain using two parameters to stimulate the probability of rainfall occurrence. Alongside, a gamma distribution with two parameters that is mean and standard deviation of wet day rainfall was used to stimulate the wet day rainfall depths. The study findings revealed that given that the parameters of the gamma distribution are unbiasedly selected from a fitted distribution, the level of variability at both short and long temporal variability can be preserved. The variability of wet periods however can be preserved if the Markov chain parameters are changed every decade.

2.3 Markov Regime-Switching Model for Modeling Time Series Data

It is not uncommon to use various time series models to analyze the dynamic behavior of variables such as rainfall, temperature, exchange rates among others. The preferred and leading models being the autoregressive (AR) models, moving

average (MA) models, and mixed autoregressive moving averages (ARMA) models.

These models have been quite successful in the field of the study. However, in many nonlinear dynamic patterns such as asymmetry, amplitude dependence and volatility clustering among others, it becomes difficult if not impossible to apply models such as the AR, MA and mixed ARMA models to analyze the dynamic behavior of the variables.

One of the nonlinear time series models that have become popular and used to characterize the behavior of different time series regimes is the Markov switching models (MSM). Regime switching models are models which can characterize time series properties in different regimes (Ayodeji, 2016). The Markov switching model according to Hamilton (1989) involves a series of models that are able to characterize the time series behavioral properties under different regimes. Under regime switching, the parameters in the model are allowed to take on different values in each of some fixed number of regimes. Application of regime switching is best suited where there is a cycling between regimes. As the regime switches from one cycle to the other, the model is able to adequately capture the dynamic behavior and properties of the distinct regimes.

A novel feature of the Markov switching model is that the switching mechanism is controlled by an unobservable state variable that follows a first-order Markov chain. In particular, the Markovian property regulates that the current value of the state variable depends on its immediate past value. As such, a structure may

prevail for a random period of time, and it will be replaced by another structure when a switching takes place (Kuan, 2002). The Markov regime switching has been applied in various fields of study to model and forecast the dynamic behavior of variables of interest. Various researchers have used different state Markov regime switching models to conduct their studies.

Gallo et al. (2015) applied a two state Markov regime-switching framework in describing El Niño Southern Oscillation (ENSO) Patterns. They observed that the pattern of behavior of the two phases (El Niño and La Niña episodes) under the study are opposite and distinguishable. The study findings also concluded that even though the Box-Jenkins methodology leads to good description of the time series of interest, it is however not able to capture certain nonlinearities that occur due to the existence of changing regimes. Added to that it was also realized that in characterizing climatic variables, one source of non-linear parameters in many meteorological time series is necessitated by the presence of weather cycles. The model used was able to capture the characteristics of the index over time. The model identified the peaks and drastic changes in the ENSO along time. Thus, the introduction of the cyclical phases of El Niño and La Niña allows characterizing in a proper patterns.

Evarest et al. (2016) modeled the temperature dynamics using two state Markov regime switching models. The two states proposed for the study is assumed to consist of a based regime controlled by a mean-reverting process with heteroskedasticity and a shifted regime also controlled by a Brownian motion. The study was also based on the foundation that a Markov regime-switching

model is not directly observable, but rather the switching process is governed by the transition probabilities. The study used current temperatures such as heating degree days (HDDs), cooling degree days (CDDs), and cumulative average temperature (CAT) which are mainly used in European cities for summer months to construct the weather derivatives. The results from the study revealed that the model developed was a good reflection of the temperature dynamics. This is because the model generated temperature indices that are approximately similar to the true temperature indices.

Chang et al. (2014) introduced a new approach to modeling regime switching using an autoregressive latent factor. This autoregressive latent factor controls the regimes depending upon whether it takes a value above or below some domain. Also, the latent factor is allowed to correlate with the innovation to the observed time series. The findings from the study revealed, given that the underlying autoregressive latent factor is stationary and independent of the model innovation, the model reduces to the conventional Markov switching model. The study concluded that the evidence of the presence of endogeneity in regime switching appears to be strong and unambiguous. It therefore suggests that neglecting endogeneity in regime switching leads to substantial bias and significant loss of information in estimating model parameters.

Gyamerah et al. (2018) conducted a study for weather derivatives using regime-switching temperature dynamics model. The study aimed at developing a time-varying mean-reversion Levy regime-switching (TML) temperature dynamics model. This model will then be able to capture the normal variations and extreme

variations in temperature which will characterize the stochastic dynamics of temperature in the Bole and Tamale enclave. The study made use of a two-state Markov regime switching model with the notion that different regimes can capture distinct principal weather conditions as well as the localized weather pattern. That is, the daily temperature is said to be latent with two possible regimes, either in the base regime (mean-reverting regime) or in the shifted regime (extreme regime). Findings from the study revealed that even though the mean-reversion rate of Tamale is higher than that of Bole, the speed of the mean-reversion rate was moderately low for both areas.

Bazzi et al. (2014) studied the time varying transition probabilities for Markov regime switching models. In their attempt to develop the model, the transition probabilities were allowed to vary over time as specific transformations of the lagged observations. The scores of the predictive likelihood function were used to specify a suitable functional form to link past observations to future transition probabilities. The model was applied in the industrial production growth as well as the financial sector. The study results showed that the model dynamic features can be used in the mean and variance simultaneously to study the industrial production growth.

Hayashi et al. (2013) developed a regime-switching structural vector autoregression (SVAR) for the analysis of quantitative easing (QE). The developed model incorporated the exit condition for terminating quantitative easing and was applied to the Japanese economy that had accumulated great reserves. The study illustrated that an increase in the reserves under quantitative

easing will lead to a raise in the inflation rate and output. Relative to the existing Japanese literature, this conclusion was reached under the constraint of taking into account the regime endogeneity. That is the effect of the quantitative easing on the inflation rate and output will be considerably smaller if regime endogeneity is taken into account.

2.4 Summary

This chapter presented a review of works on rainfall modeling forecast and Markov regime-switching model. Albeit, different researchers have used various forms of models to model and forecast rainfall in the Upper East Region, these models have been inadequate to describe the rainfall. This is especially with the switching of the rainfall seasons. This study therefore uses the Markov regime-switching model to model and estimate the expected rainfall regimes in the Upper East Region.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

This chapter presents the methods that will be used to achieve the objectives of the study. The chapter presents the description of the study area, source of data, the theoretical framework on Markov regime switching autoregressive models and the empirical framework of the study.

3.2 Profile of the Study Area

The Upper East region is located in the north-eastern enclave of Ghana. The region is bordered by Burkina Faso to the north, Togo to the east, North East Region to the south and Upper West Region to the west. The region is located approximately between longitude 0° and 1° west, and latitudes 10° 30′N and 11°N. The Upper East region is split into 15 districts or municipal, each headed by a district or municipal chief executive. The Upper East Region is one of the sixteen regions of Ghana and is the second smallest of all the regions in the country with a complete land surface of 8,842 square kilometers representing about 2.7 per cent of the total land area of Ghana. The region has its regional capital at Bolgatanga.

The upper east region according to the 2010 population and housing census has a total population of 1,046,545, which constitutes 4.2% of the nation's population. Upper east region is the least urbanized among the regions with a rural population of over 79%. The region has an active population of 84.6% thus 15 years and older

(GSS, 2012). The Upper East Region is part of the interior continental climatic zone of the country characterized by pronounced dry and wet seasons. The two seasons (dry season and rainy season) are influenced by two oscillating air masses. First is the warm, dusty, and dry harmattan air mass (dry season), which blows from the northeasterly direction across the whole municipality from the Sahara Desert. During the period of its influence (late November to early March), the zone does not experience any form of rainfall. Also, the vapor level and the vapor pressure level are usually very low, which is less than 10mb with relative humidity rarely exceeding 20 percent during the day but may rise to 60 percent during the nights and early mornings. Temperatures are usually moderate at this time of the year, ranging from to base on meteorological service records. The rainy season (wet season) starts from April and ends in October. During this period, the whole of the West African sub-region including the Upper East Region is under the influence of a deep tropical maritime air mass. This air mass together with rising convectional currents provides the Region with rains. The zone receives an annual average rainfall of about 800 mm with extreme variability and reliability both between and within the seasons. Another striking characteristic is the large quantity of rainwater normally lost through evapotranspiration from open water surfaces. Estimates of the volume of rainwater loss vary from 1.55 mm 1.65 mm per annum (GSS 2012). The industrial activity is predominantly agriculture, with 46.5 percent engaged in the activity. The closest to this form of economic activity is wholesale and retail; repair of motor vehicles and motorcycles which constitute 21.7 percent (GSS 2012).

3.3 Source of Data

To achieve the objectives of the study, secondary data on monthly rainfall is collected from the Ghana Meteorological Agency. The data is obtained from two meteorological stations at Bolgatanta, and Navrongo weather substations, all located in the Upper East region. The data covers the period year 1999 to year 2019.

The first step that regulates time series model is to investigate whether the series is stationary or otherwise. Performing a unit root test for the data is important to determine whether the data is stationary or otherwise.

3.4 Data Exploration

The data sets are examined for missing data. Where missing data is identified, the method of combined average is used to fill in the missing data. The combined average is computed using two distinct averages. In this study, the averages of seven months or years before and after the missing data time period are employed. Monthly rainfall data is ideal for the study in that the distribution of the rainfall in the study areas is easily observed on a monthly basis. Also, the monthly data provides enough data points for the time series analysis and modeling and sufficient duration for the forecast. Equation (3.1) summarizes the model in estimating the missing values.

$$T_{month}(t) = \frac{\sum_{m=1}^{7} T(t-m) + \sum_{m=1}^{7} T(t+m)}{14} \qquad (3.1)$$

3.5 Correlogram

To successfully model the rainfall data, obtaining a lag order which will give an optimal autoregressive model for modeling the data is crucial. The ACF gives values of the auto-correlation of any series with its lagged values. These lag values are plotted along with the confidence band. The ACF describes how well the present value of the series is related with its past values. Since a time series can contain components such as trend, seasonality, cyclic and residual, the ACF considers all these components while finding correlations. This therefore makes the ACF a 'complete auto-correlation plot'. The PACF instead of finding correlations of present with lags like ACF, it finds correlation of the residuals (which remains after removing the effects which are already explained by the earlier lag(s)) with the next lag value hence 'partial' and not 'complete'. Given that there is relevant information in the residual which can be modeled by the next lag, we might get a good correlation and we will keep that next lag as a feature while modeling. In modeling the series, few features are retained to avoid the issue of multicollinearity. The optimum features or order of the AR process is determined using a PACF plot, as it excludes the variations explained by earlier lags so that only the relevant features are retained.

3.6 Unit Root Test

Unit root tests also referred to as a difference stationary process are tests for stationarity in a time series analyses. A time series is stationary if a shift in time does not cause a change in the shape of the distribution (Zivot & Wang, 2006). The presence of unit roots is one cause for non-stationarity.

Autoregressive unit root tests are based on testing the null hypothesis that $\phi = 1$ (difference stationary) against the alternative hypothesis that $\phi < 1$ (trend stationary). They are called unit root tests because under the null hypothesis the autoregressive polynomial of $y_t, \phi(z) = (1 - \phi z) = 0$ has a root equal to unity.

The techniques to be employed to test for the unit root include the Augmented Dickey-Fuller test, the Kwiatkowski Phillips Schmidt–Shin (KPSS) tests, and the Philip-Perron test among others.

3.6.1 Augmented Dickey-Fuller Test

Augmented Dickey–Fuller test (ADF Test) which is an augmented version of the Dickey-Fuller (DF) test is used for a larger and more complicated set of time series models to test for the presence of unit root (Zivot & Wang, 2006). The test procedure for the ADF test is the same as the Dickey-Fuller test, only that the ADF test is applied to large and complicated set of time series models.

ADF test widen the scope of the Dickey Fuller test equation to include high order of regressive process in the model.

The Dickey-Fuller test is testing if $\phi = 0$ in the model of the data:

$$y_t = \alpha + \beta t + \phi y_{t-1} + \varepsilon_t, \tag{3.2}$$

which is written as

$$\Delta y_t = y_t - y_{t-1} = \alpha + \beta t + \gamma y_{t-1} + \varepsilon_t, \tag{3.3}$$

where y_t is the data for the study. It is written this way so that a linear regression of Δy_t against t and y_{t-1}, and test if $\gamma \neq 0$. If $\gamma = 0$, then we have a random walk process. If not and $-1 < 1 + \gamma < 1$, then we have a stationary process.

The Augmented Dickey-Fuller test allows for higher-order autoregressive processes by including Δy_{t-p} in the model. But the test under the study is still if $\gamma = 0$. The series can be derived by employing the lagged-differenced equation as stated in the model equation (3.4).

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \ldots + \delta_p \Delta y_{-p} + \varepsilon_t, \qquad (3.4)$$

where $\Delta y_t = y_t - y_{t-1}$ for y_t as data at time t, δ_i is the coefficient of the autoregressive term, γ is the coefficient of the observed data at t and ε_t represents the residual term at time t.

To investigate the stationary behavior of the data, the ADF test hypothesis is stated as;

$H_0 : \gamma = 0$ (data is none stationary)

$H_0 : \gamma \neq 0$ (data is stationary)

where the ADF test statistic is estimated using the model (3.5)

$$ADF = \frac{\gamma}{SE(\gamma)}, \qquad (3.5)$$

where γ is the mean estimate of the parameter γ for observed data and $SE(\gamma)$ is the standard error of the mean estimate of the parameter γ.

The null hypothesis of the tests is stated as the data is non-stationary. The null hypothesis will be rejected for a *p*-value less than 0.05.

3.6.2 Philip-Perron Test

The Phillips-Perron (PP) unit root tests differ from the ADF tests mainly in how the PP deal with serial correlation and heteroskedasticity in the errors. In particular, where the ADF tests use a parametric autoregression to approximate the ARMA structure of the errors in the test regression, the PP tests ignore any serial correlation in the test regression (Zivot & Wang, 2006). The test regression for the PP tests is

$$\Delta y_t = \beta' D_t + \pi y_{t-1} + u_t, \tag{3.6}$$

where u_t is $I(0)$ and may be heteroskedastic. The PP tests correct for any serial correlation and heteroskedasticity in the errors u_t of the test regression by directly modifying the test statistics $t_\pi = 0$ and $T\pi$. These modified statistics, denoted Z_t and Z_π, are given by

$$y_t = \left(\frac{\sigma^2}{\lambda^2}\right)^{\frac{1}{2}} .t_{\pi=0} - \frac{1}{2}\left(\frac{\lambda^2 - \sigma^2}{\lambda^2}\right).\left(\frac{T.SE(\pi)}{\sigma^2}\right). \tag{3.7}$$

$$y_\pi = T\pi - \frac{1}{2}\frac{T^2.SE(\pi)}{\sigma^2}\left(\lambda^2 - \sigma^2\right). \tag{3.8}$$

The terms σ^2 and λ^2 are consistent estimates of the variance parameters

$$\lambda^2 = \lim_{T\to\infty} \sum_{t=1}^{T} E[T^{-1} S_T^2], \tag{3.9}$$

$$\sigma^2 = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E[u_t^2], \tag{3.10}$$

where $S_T = \sum_{t=1}^{T} u_t$.

Under the null hypothesis that $\pi = 0$, the PP y_t and y_π statistics have the same asymptotic distributions as the ADF t-statistic and normalized bias statistics. One advantage of the PP tests over the ADF tests is that the PP tests are robust to general forms of heteroskedasticity in the error term u_t. Also, the lag length for the test regression need not be specified for a PP test.

3.6.3 Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Test

Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests are used for testing a null hypothesis that an observable time series is stationary around a deterministic trend against the alternative of non-stationarity (Zivot & Wang, 2006). The most commonly used stationarity test, the KPSS test, is due to Kwiatkowski, Phillips, Schmidt and Shin (1992). They derive their test by starting with the model

$$yt = \beta' D_t + \mu_t + u_t. \tag{3.11}$$

$$\mu_t = \mu_{t-1} + \varepsilon_t, \varepsilon_t \sim WN(0, \sigma_\varepsilon^2), \tag{3.12}$$

where D_t contains deterministic components (constant or constant plus time trend), u_t is $I(0)$ and may be heteroskedastic. Notice that μ_t is a pure random walk with innovation variance σ_ε^2 The null hypothesis that y_t is $I(0)$ is formulated as $H_0 : \sigma_\varepsilon^2 = 0$, which implies that μ_t is a constant. The KPSS test statistic is the Lagrange multiplier (LM) or score statistic for testing $\sigma_\varepsilon^2 = 0$ against the alternative that $\sigma_\varepsilon^2 > 0$ and is given by

$$KPSS = \left(T^{-2} \sum_{t=1}^{T} S_t^2 \right) / \lambda^2, \tag{3.13}$$

where $S_t = \sum_{j-1}^{t} u_j, u_t$ is the residual of a regression of y_t on D_t and λ^2 is a consistent estimate of the long-run variance of u_t using u_t .Under the null that y_t is $I(0)$, I(0), Kwiatkowski, Phillips, Schmidt and Shin show that KPSS converges to a function of standard Brownian motion that depends on the form of the deterministic terms D_t but not their coefficient values β In particular, if $D_t = 1$ then

$$KPSS \xrightarrow{d} \int_0^1 V_1(r) dr, \tag{3.14}$$

where $V_1(r) = W(r) - rW(1)$ and $W(r)$ is a standard Brownian motion for $r \in [0,1]$. If $D_t = (1,t)'$ then

$$KPSS \xrightarrow{d} \int_0^1 V_1(r)dr, \tag{3.15}$$

where $V_2(r) = W(r) + r(2-3r)W(1) + 6r(r^2-1)\int_0^1 W(s)ds$. The stationary test is a one-sided right-tailed test so that one rejects the null hypothesis of stationarity at the $100.\alpha\%$ level if the KPSS test statistic (3.14) is greater than the $100(\alpha-1)\%$ quartile from the appropriate asymptotic distribution (3.15)

3.7 Markov Switching Autoregressive (MSAR) Model Analysis

The Markov chain constitutes the reference point of the MSAR model. The Markov chain is an approach that describes how the probability of an event depends only on the state attained in the previous event. A Markov switching model is modeled by combining two or more dynamic models through a Markovian switching mechanism.

A Markov switching model is modeled by combining two or more dynamic models through a Markovian switching mechanism. Following Hamilton (1989, 1994), a simple Markov switching model is stated as below.

3.7.1 Markov Switching (MS) Model

A Markov switching (MS) is pair of discrete-time stochastic processes with one process being observed and the other being unobserved which is sometimes referred to as hidden (latent). The dynamics of the observed process is determined by the dynamics of the unobserved process. This makes it possible for the distribution of the unobserved process to be reconstructed by the series of observations (Kuan, 2002).

The latent process is a finite-state Markov chain, defined as S_t whereas the observed process is defined as Y_t, provided that the Markov chain is conditionally autoregressive. Given this theoretical structure, an MSAR model permits the following;

a. Modeling nonlinear and non-normal time series by holding the assuming that different autoregressions, each one depending on a hidden state, alternate according to the Markovian regime switching.
b. Classifying the observations into a small number of homogeneous groups, identified as the regimes of the Markov chain.

Let S_t denote an unobservable state variables assuming the value one or two. A simple switching model for the variable Y_t involves two AR (1) specifications:

$$Y_t = \begin{cases} \alpha_0 + \beta y_{t-1} + \varepsilon_t, S_t = 1 \\ \alpha_0 + \alpha_1 + \beta y_{t-1} + \varepsilon_t, S_t = 2 \end{cases}, \quad (3.16)$$

where, $|\beta| < 1$ and ε_t are independent and identically distributed (i.i.d) random variables with mean of zero and variance σ_ε^2. This is a stationary AR (1) process with mean of $\alpha_0/(1-\beta)$ given that $S_t = 1$. The process then switches to another stationary AR (1) process with mean $(\alpha_0 + \alpha_1)/(1-\beta)$ when S_t switches from 1 to 2. Provided that $\alpha_1 \neq 1$, this model will assume two dynamic structures at different levels, depending on the value of the state variable S_t. In this case, y_t

are governed by two distributions with distinct means, and S_t determines the switching between these two distributions (regimes).

3.7.2 Autoregressive (AR) Models

An AR model is a time series model in which a value from a time series is regressed on previous values from that same time series data. For example, Y_t on Y_{t-1}. For instance

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t. \tag{3.17}$$

In the model given by (3.17), the response variable in the previous time period $\{y_{t-1}\}$ has become the predictor in the model. The errors $\{\varepsilon_t\}$ have the usual assumptions, thus independent and identically distributed and have an expectation of zero. The order of an autoregression depends on the number of immediately past values in the series that are used to predict the value at the present time. The model given by (3.17) is a first-order autoregression, written as AR (1).

A second-order autoregression, written as AR (2) has the two past observations as the predictor in the model. This can be written as;

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t. \tag{3.18}$$

Generally, the p^{th}-order autoregression, written as AR (p) is a multiple linear regression in which the value of the series at any time t is a (linear) function of the previous values at times $t-1, t-2, ..., t-p$. This model can be written as below;

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + ... + \alpha_p y_{t-p} + \varepsilon_t. \quad (3.19)$$

In model given by equation (3.19), $\{y_t\}$ is predicted using the previous series up to time $t-p$.

3.7.3 Markov Switching Autoregressive Models

A Markov switching autoregressive (MSAR) process is a discrete-time process with two components $\{S_t, Y_t\}$. For our particular application, $\{Y_t\}$ denotes the observed process with values in $(1, +\infty)$ and $S_t \in (1, ..., M)$ represents the latent or hidden process at time t (Ailliot and Monbet, 2011).

Given that the likelihood of a regime S_t is equivalent to a particular value of j relies solely on the past value of S_{t-1}, then a Markov chain of this kind can be specified in the model (3.20).

$$P(s_t = j \mid S_{t-1} = i, S_{t-1} = k, ...) = P(s_t = j \mid S_{t-1} = i) = pij, \quad (3.20)$$

where p_{ij} is referred to the transition probability that satisfies the model equation in (3.21).

$$p_{i1} + p_{i2} + ... + p_{iM} = 1. \quad (3.21)$$

The transition probability p_{ij} illustrates the chance of transition from regime j after the event of regime i.

The MSAR function governs a basic variation with an unobserved regime that satisfies the first order Markov chain. Markov properties define the estimate of the basic variable given the estimate of the previous estimate. The general form of the MSAR model is summarized in equation (3.22).

$$(y_t - \mu_{s_t}) = \sum_{p=1}^{M} \delta_p (y_{t-p} - \mu_{s_{t-p}}) + \varepsilon_t, \tag{3.22}$$

where y_t refers to the observed data, δ_p is the coefficient of the autoregressive component, S_t is the state at time t, the term μ_{s_t} is a constant and depends on the state S_t and residual ε_t at time t.

The likelihood density model of y_t given the stochastic variable S_t as the conditional likelihood density model can be represented in the model (3.23).

$$f(y_t \mid s_t = j; \mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi\sigma_j}} \exp\left(\frac{-(y_t - \mu_j)^2}{2\sigma_j^2} \right), \tag{3.23}$$

for all $j = 1, 2, ..., M$.

The unexpected regime $\{S_t\}$ is determined by certain stochastic distribution. For the marginal likelihood of S_t and j, the model is represented by π_j as illustrated in equation (3.24).

$$P(S_t = j; \theta) = \pi_j, \tag{3.24}$$

given that $j = 1,2,...,M$. the likelihood of $\pi_1, \pi_2, ..., \pi_M$ is inclusive of θ, where $\theta = (\mu_1, \mu_2, ..., \mu_M, \sigma_1^2, \sigma_2^2, ..., \sigma_M^2, \pi_1, \pi_2, ..., \pi_M)$. The likelihood of mutual occurence $S_t = j$ and y_t on the interval [m,n] can be determined by

$$P(Y_t, S_t = j; \theta) = f(y_t \mid s_t = j; \mu_j, \sigma_j^2) P(s_t = j; \theta)$$
$$= \frac{1}{\sqrt{2\pi\sigma_j}} \exp\left(\frac{-(y_t - \mu_j)^2}{2\sigma_j^2} \right). \quad (3.25)$$

The marginal likelihood density model of y_t is then estimated by obtaining the sum in the model (3.25) for every possible value of j. This yields the function below

$$f(y_t; \theta) = \sum_{j=1}^{M} P(y_t, s_t = j; \theta). \quad (3.26)$$

3.8 Parameter Estimation

The parameters of the model can be estimated by employing Maximum Likelihood Estimation (MLE) method. In doing so, the probability density function is first and foremost estimated. The log-likelihood function thereof can be derived from the probability density function. With reference to the transition probability condition in the model (3.21) and the MSAR model in (3.22), the density function can be obtained from a basic analysis of two regime Markov switching model with an autoregressive term of order one, as in model equation (3.27).

$$f(y_t \mid s_t, s_{t-1}, \Gamma_{t-1}; \theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{((z_t - \mu_{s_t}) - \delta_1 (y_{t-1} - \mu_{s_t}))^2}{2\sigma^2} \right). \quad (3.27)$$

where $\Gamma_{t-1} = (y_{t-1}, y_{t-2}, ...)$ is the population data and $\theta = (\mu_1, \mu_2, \sigma^2, \delta_1)$ is pool of parameters of the modekl MS(2)-AR(1). The likelihood density model Y_t is derived from estimating the density function together for all possible values of S_t and S_{t-1} as follows

$$\begin{aligned} f(y_t \mid s_t, s_{t-1}, \Gamma_{t-1}; \theta) &= \sum_{j=1}^{M} \sum_{i=1}^{M} f(y_t, s_t = j, s_{t-1} = i \mid \Gamma_{t-1}; \theta) \\ == \sum_{j=1}^{M} \sum_{i=1}^{M} f(y_t, s_t = j, s_{t-1} = i \mid \Gamma_{t-1}; \theta) &\times P(s_t = j, s_{t-1} = i \mid \Gamma_{t-1}; \theta). \end{aligned} \quad (3.28)$$

The filtering and smoothing process is used to determine the likelidood of regime of observation t from past observation to observation $t-1$. The filtering process is carried out to determine the state at time t with reference to the observational data until time t. The model in determining the filtering process is as specified in the equation (3.29) as follows

$$P(s_t = j, s_{t-1} = i \mid \Gamma_{t-1}; \theta) = \frac{f(y_t \mid s_t = j, s_{t-1} = i \mid \Gamma_{t-1}; \theta) P(s_t = j, s_{t-1} = i \mid \Gamma_{t-1}; \theta)}{\sum_{j=1}^{M} \sum_{i=1}^{M} f(y_t \mid s_t, s_{t-1} \mid \Gamma_{t-1}; \theta) P(s_t = j, s_{t-1} = i \mid \Gamma_{t-1}; \theta)}. \quad (3.29)$$

The filtered regime likelihood value of a regime can be computed by using defined model equation stated in (3.30).

$$P(s_t = j, s_{t-1} = i \mid \Gamma_{t-1}; \theta) = \sum_{i=1}^{M} P(s_t = j, s_{t-1} = i \mid y_t, \Gamma_{t-1}; \theta). \quad (3.30)$$

The smoothing process is determined using information of the observational data until $t = T$ to characterize the likelihood of the regime value. This is done to obtain the estimation value. The equation for the smoothing process is as in model equation (3.31).

$$P(s_t = j, s_{t+1} = k \mid \Gamma_{t-1}; \theta) = \frac{Ps_{t+1}, \Gamma_T Ps_t \Gamma_t Ps_{t+1}, \Gamma_T}{P(s_{t+1} = k \mid \Gamma_t; \theta)}, \quad (3.31)$$

Where

$$\begin{aligned} Ps_{t+1}, \Gamma_T &= P(s_{t+1} = k \mid \Gamma_T; \theta) \\ Ps_t, \Gamma_t &= P(s_t = j \mid \Gamma_t; \theta) \\ Ps_{t+1}, \Gamma_t &= P(s_{t+1} = k \mid s_t = j, \Gamma_t; \theta). \end{aligned} \quad (3.32)$$

The equation (3.32) is calculated for every possible value of k. The probability value of S_t with value j based on observations until time $t = T$ is determined as

$$P(s_t = j \mid \Gamma_T; \theta) = \sum_{k=1}^{M} P(s_t = j, s_{t+1} = k \mid \Gamma_T; \theta). \quad (3.33)$$

This model has the capability to determine the likelihood of y_t for all values of s_t amidst the course of filtering and smoothing. The probability density function of Y_t can be obtained as follows

$$\begin{aligned} & f(y_t \mid s_t, s_{t-1}, \Gamma_{t-1}; \theta) \\ &= \sum_{j=1}^{M} \sum_{i=1}^{M} f(y_t \mid s_t, s_{t-1}, \Gamma_{t-1}; \theta) \times P(s_t = j, s_{t-1} = i \mid \Gamma_{t-1}; \theta). \end{aligned} \quad (3.34)$$

The likelihood function is given in the model equation below

$$L(\theta) = \prod_{t=1}^{T} f(y_t \mid s_t, s_{t-1}, \Gamma_{t-1}; \theta). \tag{3.35}$$

The log-likelihood function is

$$\ln L(\theta) = \sum_{t=1}^{M} f(y_t \mid s_t, s_{t-1}, \Gamma_{t-1}; \theta). \tag{3.36}$$

For two regime model of MSAR, there are two limiting models; π_1 and π_2. The log-likelihood method is used to estimate the maximum likelihood in the Lagrange model as specified in model (3.37).

$$J(\theta) = \ln L(\theta) + \lambda(1 - \pi_1 - \pi_2). \tag{3.37}$$

To maximization of the parameters estimates on $\theta = (\mu_j, \sigma_j^2, \pi_j, \delta_j)$ is obtained by differencing each of the parameters in θ and equating to zero. This is written as below

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \sum_{t=1}^{T} \frac{1}{f(y_t; \theta)} \frac{\partial f(y_t; \theta)}{\partial \theta}. \tag{3.38}$$

The estimated parameters of MSAR can be obtained by the model equations below

$$\mu_j = \frac{\sum_{t=1}^{T} y_t P(s_t = j \mid Y_t; \theta)}{\sum_{t=1}^{T} P(s_t = j \mid Y_t; \theta)}. \tag{3.39}$$

$$\sigma_j^2 = \frac{\sum_{t=1}^{T}(y_t - \mu_j)^2 P(s_t = j \mid y_t;\theta)}{\sum_{t=1}^{T} P(s_t = j \mid y_t;\theta)}. \tag{3.40}$$

$$\pi_j = \frac{1}{T}\sum_{t=1}^{T} P(s_t = j \mid y_t;\theta). \tag{3.41}$$

$$\delta_\gamma = \frac{\sum_{t=1}^{T}\left(\sum_{j=1}^{M=1}(y_t - \mu_j)P(s_t = j \mid \Gamma_T;\theta)\right)}{\sum_{t=1}^{T}\left(\sum_{j=1}^{M=2} P(s_t = j \mid \Gamma_T;\theta)\right)}. \tag{3.42}$$

The longevity of each regime can be determined based on the realization of each regime. This realization can be obtained using the model equation given in (3.43) below.

$$E(D) = \frac{1}{1 - p_{ij}}. \tag{3.43}$$

3.9 Model Selection

The optimal model is determined from the analysis of the values of Akaike's Information Criterion (AIC), Bayesian information criterion (BIC) and Hannan and Quinn criterion (HQC). The optimal model is deemed to be the one with the lowest value among these criteria values. These models are discussed below.

3.9.1 Akaike Information Criteria (AIC)

Let K be the number of estimated parameters in a supposed statistical model. Let L be the maximum value of the likelihood function for the supposed statistical

model. Based on the above, the AIC value of the model can be determined as follows.

$$AIC = 2k - 2\ln(L). \tag{3.44}$$

Given a finite set of available models, the preferred model is the one with the minimum AIC value (Akaike 1974). Thus, AIC rewards goodness of fit as determined by the likelihood function. The AIC however include the penalty that the AIC is an increasing function of the number of estimated parameters. This is one advantage of the AIC as increasing the number of parameters in the model almost always improves the goodness of the fit (Akaike 1985).

3.9.2 Bayesian Information Criterion (BIC)

The Bayesian information criterion (BIC) also known as Schwarz information criterion is used for model selection among a finite set of models. The model with the lowest BIC is preferred. It is based, in part, on the likelihood function and somewhat linked to the AIC (*Romeynet al., 2012)*. The BIC is characterized as below.

$$BIC = k\ln(n) - 2\ln(L), \tag{3.45}$$

where

L is the maximum value of the likelihood function of the model M, that is

$L = p(x \mid \theta, M)$ where θ are the parameter values that maximizes the likelihood function;

x is the observed data set;

n is the number of data points in x , the number of observations;

k is the number of parameters estimated by the model.

3.9.3 Hannan–Quinn Information Criterion (HQC)

In statistics, the Hannan–Quinn information criterion (HQC) is a criterion for model selection. It should be noted that the HQC can be used as an alternative to the Akaike information criterion (AIC) and Bayesian information criteria (BIC). It is given as

$$HQC = -2L_{\max} + 2k\ln(\ln(n)), \tag{3.46}$$

where $L_{\max}$ is the log-likelihood, k is the number of parameters and n represents the number of observations.

3.10 Summary

Markov regime-switching autoregression models provide an analytical framework to analyze both shifts in regimes and the differential impact of explanatory variables across regimes (or states). This study sort to model and forecast the rainfall pattern and distribution. The study employed various tools of testing for the unit roots and model selection. The unit root test employed techniques such as the ADF test, the Philip-Perron test and the KPSS test. The model selection included the AIC, the BIC and the HQC. The parameter was done using the likelihood and log-likelihood function. An M-state Markov switching chain was used to determine the switching regimes. The study employed the two state Markov regime-switching autoregressive model to model the rainfall in the area.

CHAPTER FOUR

RESULTS AND DISCUSSIONS

4.1 Introduction

This chapter presents the main results and discussions of the study. Firstly, it presents the preliminary analyses of the data. These include the descriptive statistics, time series plots, and the test for unit roots in the data set of the weather substations. Secondly, the chapter presents further analysis of the data. This includes fitting of the MS-AR model for the weather stations.

4.2 Descriptive Statistics

The study uses monthly rainfall data measured in millimeters at the Bolgatanga and Navrongo weather substations from January 1999 to December 2019. Table 4.1 shows the descriptive statistics of the data sets for the Bolgatanga and the Navrongo weather substations. From Table 4.1, it can be observed that the minimum rainfall recorded for all the stations was 0.00 mm. The maximum rainfall recorded was 455.50 mm and 476.90 mm for Navrongo and Bolgatanga weather substations respectively. The mean rainfall is obtained as 84.37 mm and 83.97 mm for Navrongo and Bolgatanga weather substations respectively. It can be observed that the skewness measures are above 0, indicating that the data sets are positively skewed. This suggests that the data sets are not symmetric. The measure of excess kurtosis for the Navrongo and Bolgatanga weather substations is 0.7296 and 1.2586 respectively. Kurtosis measures the peakness of the probability distribution of a real valued random variable. The kurtosis of any

normally distributed data is approximately 3 (Jones and Grill, 1998). Given the excess kurtosis, the shape of the rainfall data for the Navrongo and the Bolgatanga weather stations are leptokurtic in nature.

Table 4.1: Descriptive Statistics for the Weather Substations

Variable	Navrongo	Bolgatanga
Minimum	0.00	0.00
Maximum	455.50	476.90
25th Quartile	0.00	0.00
75th Quartile	143.03	142.45
Mean	84.37	83.97
Median	41.15	45.80
SE Mean	6.23	6.18
SD	98.86	98.12
Skewness	1.16	1.24
Excess Kurtosis	0.73	1.26

Figure 4.1 shows the time series plots of the data sets. The figure indicates that the rainfall rises and falls in magnitude at particular times of the year. This behavior of the data indicates a nonlinear property in the series. The presence of the switch from one state of rainfall to another reveals that there are two possible states in the data sets. These states are deemed to be the occurrence of rainy season (high rainfall) and dry season (low or no rainfall).

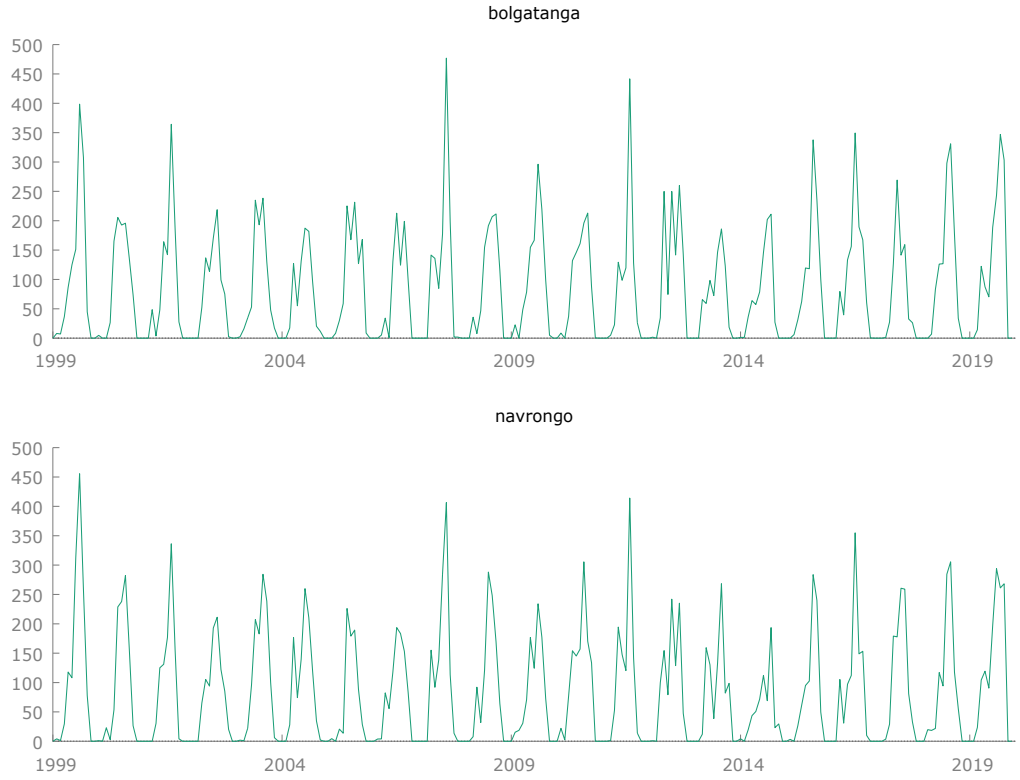

Figure 4.1: Time Series Plots of Rainfall Data for the Substations

To understand the distribution of the rainfall pattern in the study area, the monthly rainfall is computed for the study period. This data is then plotted to depict the average rainfall received per month for the study period. This is shown in Figure 4.2. The figure illustrates the states and concentration of rainfall as identified in Figure 4.1. The figure shows the concentration of rainfall along the months of April, May, June, July, August, September and October. The month of January, February, March, November and December experience low or no rainfall. It can be observed that in all the weather substations, for the month of January, February, November and December the area experiences virtually no amount of rainfall. The amount of rainfall received begins to increase gradually in quantity from the month of March and peaks in the month of August. The rainfall then decreases

thereafter and continues to decline until the month of November and December where no amount of rainfall is received.

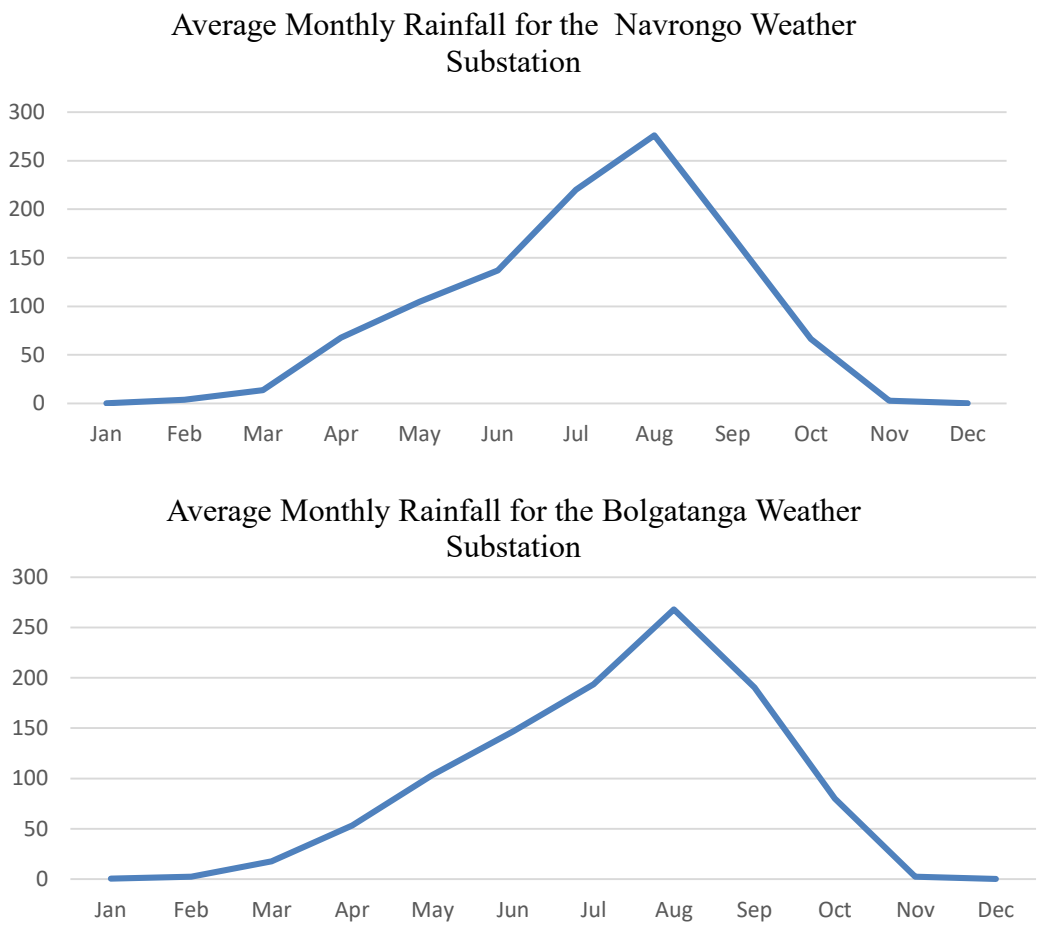

Figure 4.2: Plots of Monthly Average Rainfall for the Substations

Table 4.2 depicts the average monthly rainfall recorded over the study period. From Table 4.2, it can be observed that in all the weather substations, for the months of January, February, November and December, there is minute or no rainfall received. The study area begins experiencing rainfall from the month of

March but in small quantity. For instance, for the month of March over the study period, the Navrongo weather substation received an average of 12.84 mm and Bolgatanga received 16.83 mm of rainfall. The amount of rainfall received in the area increases gradually and peaks in the month of August. The Navrongo weather substation receives an average of 262.77 mm and Bolgatanga receives 255.32 mm of rainfall for the month of August over the study period. The amounts of rainfall received begin to decline in quantity after August until the month of December where virtually no rainfall is received.

Table 4.2: Monthly Average Rainfall for the Weather Substations

Month	Navrongo	Bolgatanga
January	0.250	0.30
February	3.60	2.12
March	12.84	16.83
April	64.40	50.69
May	99.70	98.52
June	130.53	139.85
July	209.41	184.40
August	262.77	255.32
September	163.30	181.19
October	63.08	76.19
November	2.76	2.29
December	0.00	0.00

4.3 Test for Stationarity of the Data

The first step that governs this time series analysis is to investigate the presence of a unit root in the series or otherwise. Stationarity test for the data is carried out

to determine the presence of a unit root or otherwise. The techniques employed here are the ADF, KPSS and the PP test. The results are as displayed in Table 4.3.

Table 4.3 summarizes the ADF test for the Navrongo and Bolgatanga weather substations. This test is carried out with various exogenous variables. That is with constant, without constant and with constant and trend. The test result showed that the Navrongo series contains a unit root irrespective of whether the exogenous variable included or otherwise. This is because, the *p*-value of all the test result is greater than 0.05, hence the null hypothesis is not rejected. The test result of the Bolgatanga series shows that the series is only stationary if a constant is included. The *p*-value of the test result with constant is less than 0.05.

Table 4.3: ADF Unit Root Test for the Weather Substations

Weather Substation	Test Statistic	P-value
Without Constant		
Navrongo	-0.46	0.52
Bolgatanga	-0.21	0.61
With Constant		
Navrongo	-2.83	0.051
Bolgatanga	-3.31	0.02
With Constant and Trend		
Navrongo	-2.63	0.27
Bolgatanga	-3.39	0.052

Employing the KPSS test, the null hypothesis is tested that the original series is stationary at the non-seasonal level of 5%. The test results as summarized in Table

4.4 indicate that the null hypothesis is rejected and hence the series for both stations contain a unit root.

Table 4.4: KPSS Unit Root Test for the Weather Substations

Weather Substation	5% critical value	LM-Stat.
Navrongo	0.15	0.17
Bolgatanga	0.15	0.37

Table 4.5 contains the PP unit root test result. The test results for the Navrongo and Bolgatanga series show that the series contain a unit root.

Table 4.5: PP Unit Root Test for the Weather Substations

Weather Substation	Test Statistic	Critical Value (5%)
Navrongo	-5.98	-2.87
Bolgatanga	-6.22	-2.87

Given that the rainfall data for both weather substations contains a unit root based on stationarity test performed, the data is differenced at various levels in order to achieve stationarity.

The ADF test on the first order difference data is displayed in Table 4.6. Considering the p-values for the two substations from Table 4.6, the data becomes stationary since the first difference produces *p*-values are less than the level of significance at 0.05.

Table 4.6: ADF Test on First Difference of Data

Weather Substation	Test-statistic	p-value
Navrongo	-14.68	<0.0001
Bolgatanga	-11.48	<0.0001

Table 4.7 summarizes the KPSS test statistics for the series. The LM statistic for both substations reveals the series becomes stationary after first transformation at 5% significance of level.

Table 4.7: KPSS Test on First Difference of Data

Weather Substation	5% critical value	LM-Stat.
Navrongo	0.46	0.06
Bolgatanga	0.46	0.05

Table 4.8 contains the first transformed PP unit root test result. Since the *p*-values for the series is less than the significance level at 5%, the null hypothesis is rejected. Hence, the test result for the Navrongo and Bolgatanga series shows that the series is free from unit root after the first difference.

Table 4.8: PP Test on First Difference of Data

Weather Substation	Test Statistic	Critical Value
Navrongo	-23.93	<0.0001
Bolgatanga	-22.36	<0.0001

The time series plots of the differenced data for the weather substations are displayed in Figure 4.3.

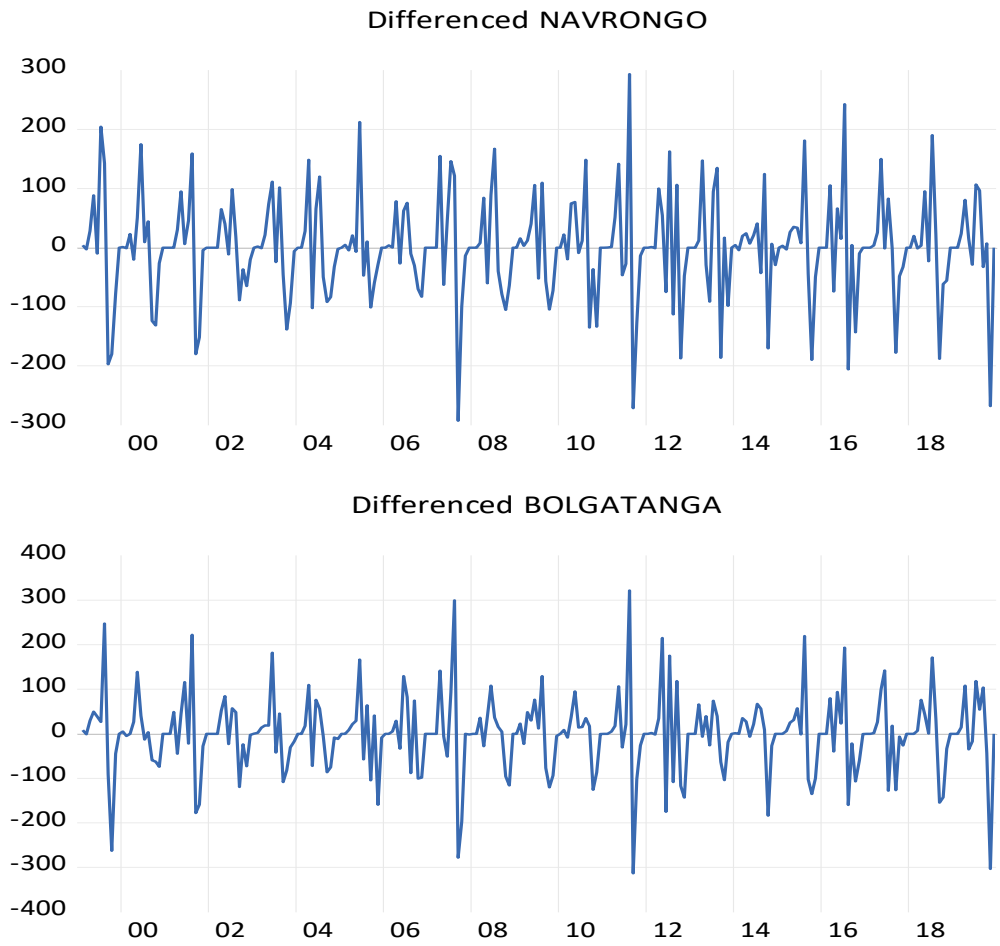

Figure 4.3: Plots of the Difference Data for the Substations

4.4 Correlogram

The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots are used to identify the number of the autoregressive terms that are significant for the estimation of the parameters including autoregressive coefficients and the transitional probabilities. To successfully model the

Navrongo rainfall data, obtaining a lag order which will give an optimal autoregressive model for modeling the data is crucial. The optimum order of the AR process is determined using a PACF plot, as it excludes the variations explained by earlier lags so that only the relevant features are retained. The other reason for using the PACF is also that it ensures that only the relevant features or lags are retained. The plots of the ACF and the PACF with lag order of 12 for the Navrongo substation are shown in Figure 4.4.

The lags that are statistically significant to be included in the model are at lag 1, lag 11 and lag 12 for the Navrongo substation.

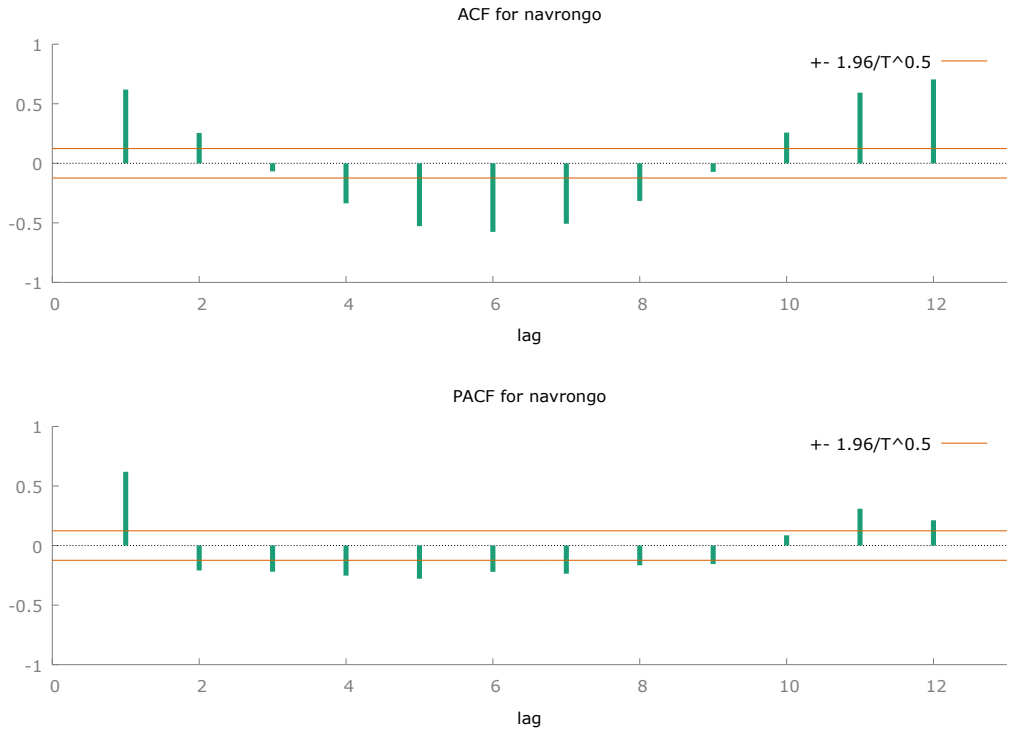

Figure 4.4: ACF and PACF Plots for the Navrongo Substation

To model the rainfall patterns in the Bolgatanga weather substation, the lag order of the substation is determined. This is done with the aid of the ACF and PACF plots. The PACF forms the basis for determining the lag order to include in modeling the. The plot of the ACF and the PACF with lag order of 12 is as in Figure 4.5. From Figure 4.5, the lags are determined based on the spikes from the PACF plot. The lags that are statistically significant to be included in the model are at lag 1 and lag 11.

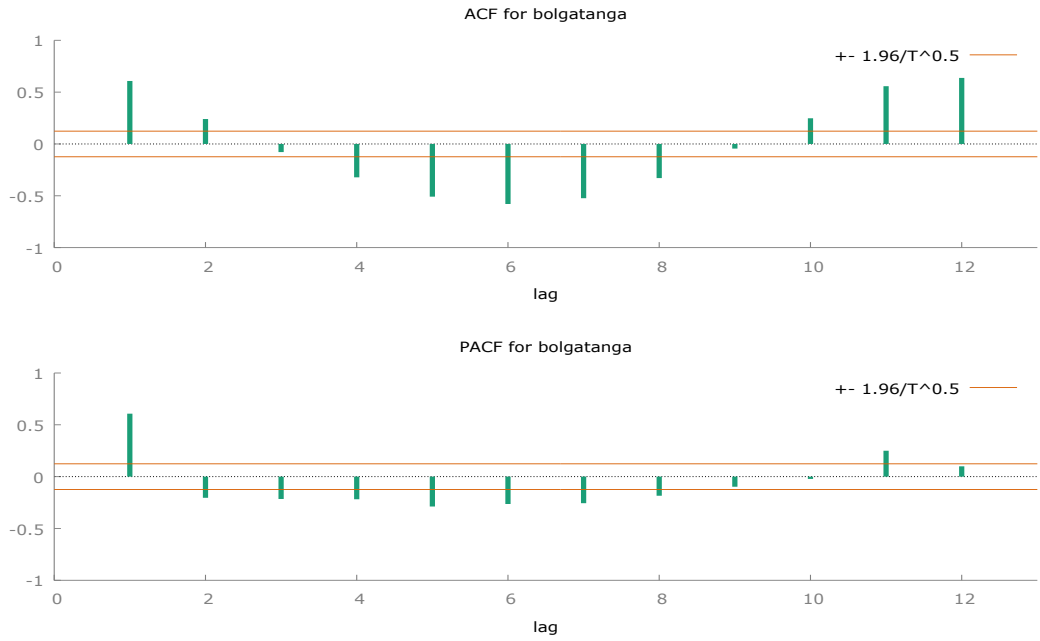

Figure 4.5: ACF and PACF plots for Bolgatanga Substation

4.5 Parameter Estimation of Markov Switching Autoregressive Model

The seasons in the study area exhibits a cyclical phenomenon such that it is possible to identify two distinct regimes (rainy season and dry season). The behavioral patterns observed during each of these two periods are opposite in nature and differentiable. What is uncertain however, is the length of the

transitional phases from one to the other. Based on the stated conditions above, the technique of the MSAR method can be employed to model the rainfall data. The rainy seasons or months with rainfall are determined to be state one (1) while the dry season or months with minimal or no rainfall are determined to be state two (2).

4.5.1 Markov Switching Autoregressive Model for Navrongo Substation

In order to model the data set with MSAR, it is imperative to establish the order of the autoregressive term p to include in capturing the dynamic nature of the phenomenon. The model is then revised to determine the adequacy of the specified and estimated model.

Based on the MS (2)-AR (p) lag selection determination, which is consistent with the PAFC plot in Figure 4.4, the AR models to consider based on the PACF plot are AR (1), AR (11) and AR (12).

The MS (2)-AR (p) models at lag 1, 11, and 12 are analyzed to determine the best model that represents the series. The AIC, BIC, and HQC values of MS (2)-AR (1), MS (2)-AR (11), and MS (2)-AR (12) are obtained. The model with the lowest AIC, BIC, and HQC ranking is deemed to outperform the other models. From Table 4.9, we obtain the optimal model using the values of the AIC, HQC, and BIC criteria to be the model MS (2) – AR (1).

Table 4.9: Model for Navrongo Substation

Model	AIC	BIC	HQC
MS (2) - AR (1)	8.92	9.00	8.95
MS (2) - AR (11)	10.96	11.05	11.00
MS (2) - AR (12)	10.56	10.65	10.60

Given the model specification, the maximum likelihood estimates of the autoregressive coefficients and the transition probabilities were obtained. The estimated parameters are summarized in Table 4.10. The parameters μ_{s_t} which is a constant and depends on the state S_t are $\mu_1 = 2.24$ and $\mu_2 = 4.29$. All of the parameters estimated are statistically significant ($p < 0.05$). Also, the parameter δ_p is the coefficient of the autoregressive component which also depends on S_t. The coefficient of the AR (1) are $\delta_1 = 0.51$ and $\delta_2 = 0.84$ for regime 1 and 2 respectively. These coefficients are statistically significant at 5% level of significance.

Table 4.10: Estimates of Regime Parameters for Navrongo Substation

Regime	Parameter	Coefficient	Std. Error	z-Statistic	*p*-value
1	μ	2.24	0.08	28.69	<0.0001
	δ	0.51	0.02	24.18	<0.0001
2	μ	4.29	0.08	54.24	<0.0001
	δ	0.84	0.11	38.30	<0.0001

The MSAR fitted model for the Navrongo weather substation can therefore be specified as;

$$y_t = \begin{cases} 0.51(y_{t-1} - 2.24), S_t = 1 \\ 0.84(y_{t-1} - 4.29), S_t = 2 \end{cases}.$$

From the fitted model, the transitional probabilities and the duration of each state are estimated. The transitional probabilities estimate the probability of moving from one state to the other after an event in the preceding state. The transition probability p_{ij} illustrates the likelihood of transition to regime j after the occurrence of regime i. These transitional probabilities are $P_{11} = 0.80$, $P_{12} = 0.20$, $P_{21} = 0.35$, and $P_{22} = 0.65$. These transitional probabilities can be written in a matrix form as;

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 0.80 & 0.20 \\ 0.35 & 0.65 \end{pmatrix}.$$

These probabilities satisfy the condition $p_{i1} + p_{i2} + ... + p_{iM} = 1$. Given that the phenomenon is in state 1, the probability of experiencing the said phenomenon remains high at approximately 80%. This explains that given a rainfall regime, the likelihood of experiencing rainfall over a month is estimated to be high at $P_{11} = 0.7963$, and the probability of rainfall is reduced to $P_{12} = 0.2037$ for periods of no rainfall or minimal rainfall. Moreover, for a period of no rainfall, the probability of no rainfall during a month is $P_{22} = 0.6486$. The probability of

experiencing rainfall in a particular period is estimated at $P_{21} = 0.3514$ given that the preceding period experienced no rainfall.

The expected duration for each regime in the period is estimated based on equation

$$E(D) = \frac{1}{1 - p_{ij}}.$$

It is estimated that the expected duration of high rainfall over the study period is at least 5.10 months while the period of no rainfall is expected to last for at least 3.05 months for every year. That is when the phenomenon is in regime one, it is expected to be in that state for 5.1 months. A switch then happens and the phenomenon is expected to stay in regime two for 3.05 months before another switch takes place. This is consistent with the historical plot of the data for the study period in Figure 4.2. The plot of the expected durations and the Markov transitional probabilities are displayed in Figure 4.6 and Figure 4.7 respectively.

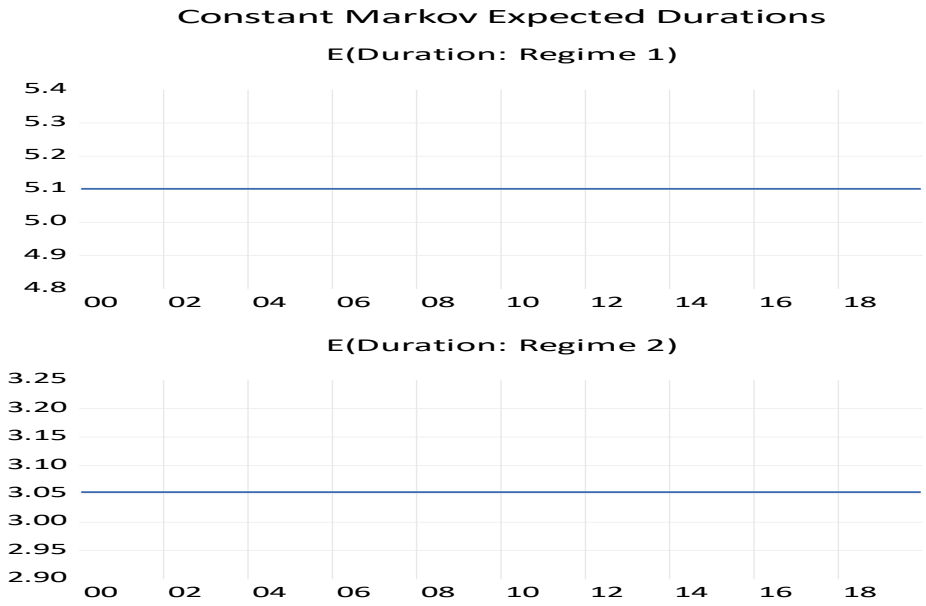

Figure 4.6: Constant Markov Expected Duration for the Regimes of Navrongo

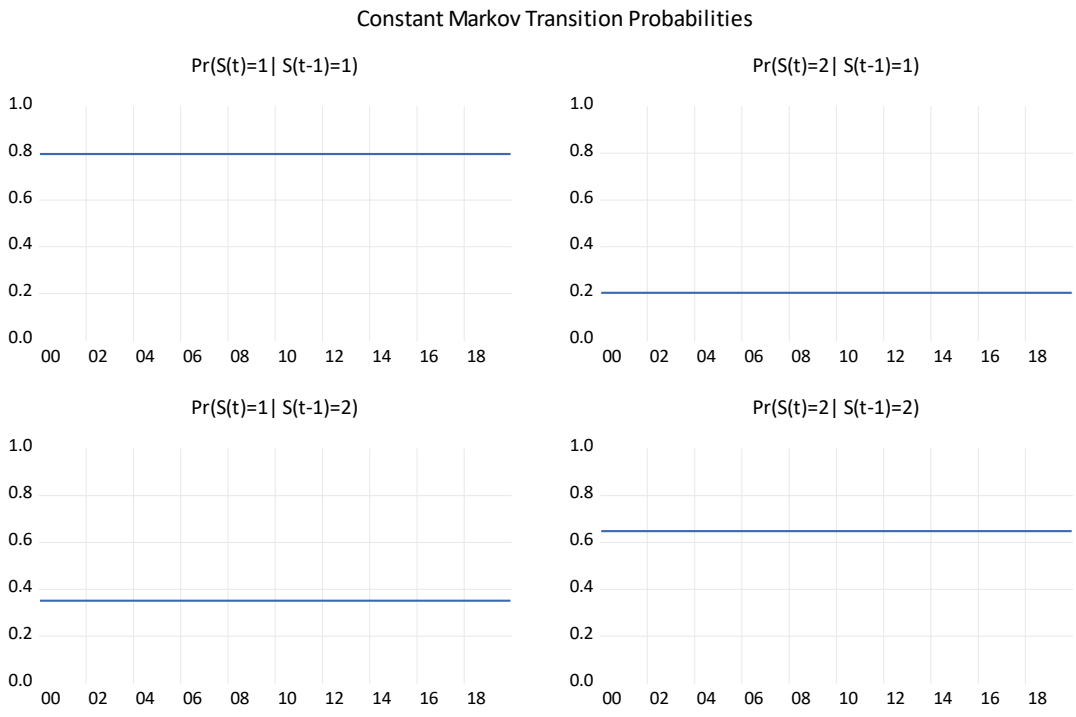

Figure 4.7: Constant Markov Transitional Probabilities of Navrongo

The smoothened probabilities are estimated with reference to the observational data until $t = T$. This is done to obtain the estimated values. To compute the smoothing probabilities $IP(s_t = i \mid Z^T; \theta)$, the method of approximation by Kim (1994) is employed. The probability value of the regimes (S_t) with value j with reference to observations until time $t = T$ is determined. This estimation has the capability to determine the likelihood of y_t for all the values of the state s_t by employing the technique of filtering and smoothing.

The plots of these filtered and smoothed probabilities are as in Figure 4.8. As it can be seen in Figure 4.8, the transition probabilities of remaining in the same state $\left(P(S(t) = 1), P(S(t) = 2\right)$ are approximated to one or zero for various seasons. Each of these periods or seasons represents a cycle of the rainy season and the dry season. During the rainy season, the probability of remaining in regime one $P(S(t) = 1)$ is close to one. This is consistent with the estimates from the transitional probabilities. Thus given a period of high rainfall, the likelihood of experiencing rainfall in the subsequent month is estimated to be $P_{11} = 0.80$.

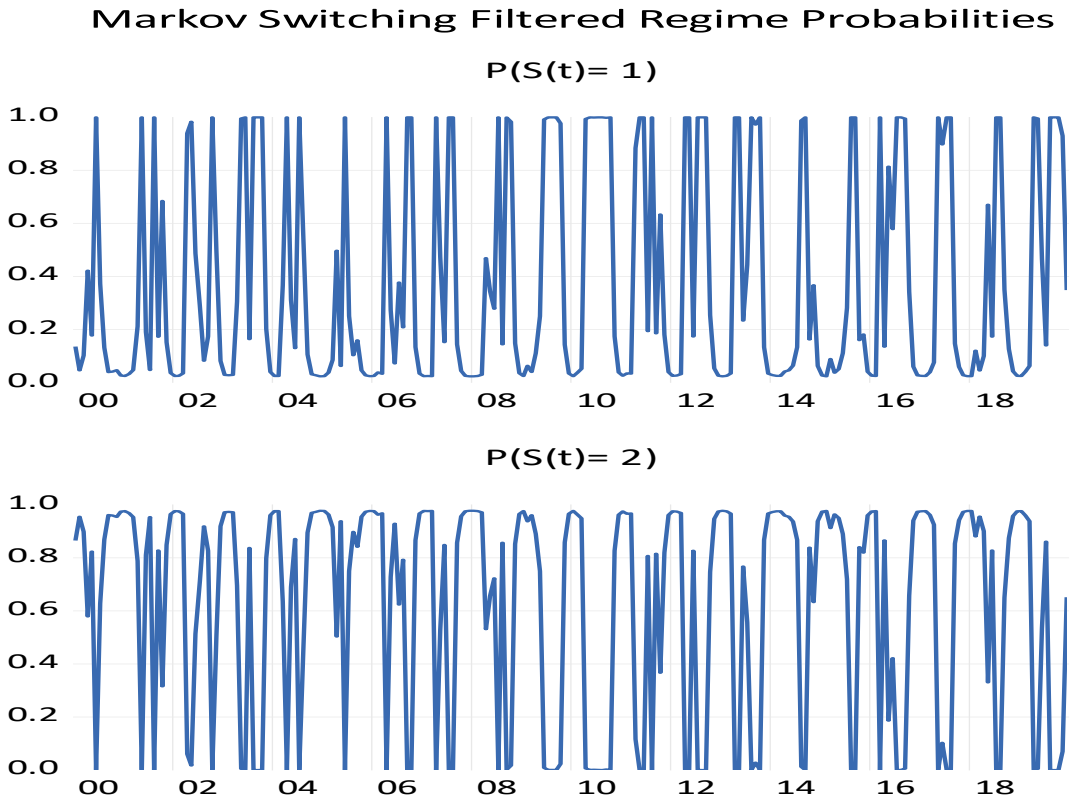

Figure 4.8: Filtered Probabilities for the Navrongo Substation

Conversely, during the dry season or period of minimal rainfall, the probability of remaining in regime 2 thus $P(S(t)=2)$ is close to zero. This thus suggests that these regimes are mutually exclusive. As it can be expected, the same pattern can be observed throughout the period of the study. Figure 4.9 illustrates the Markov switching smoothed regime probabilities.

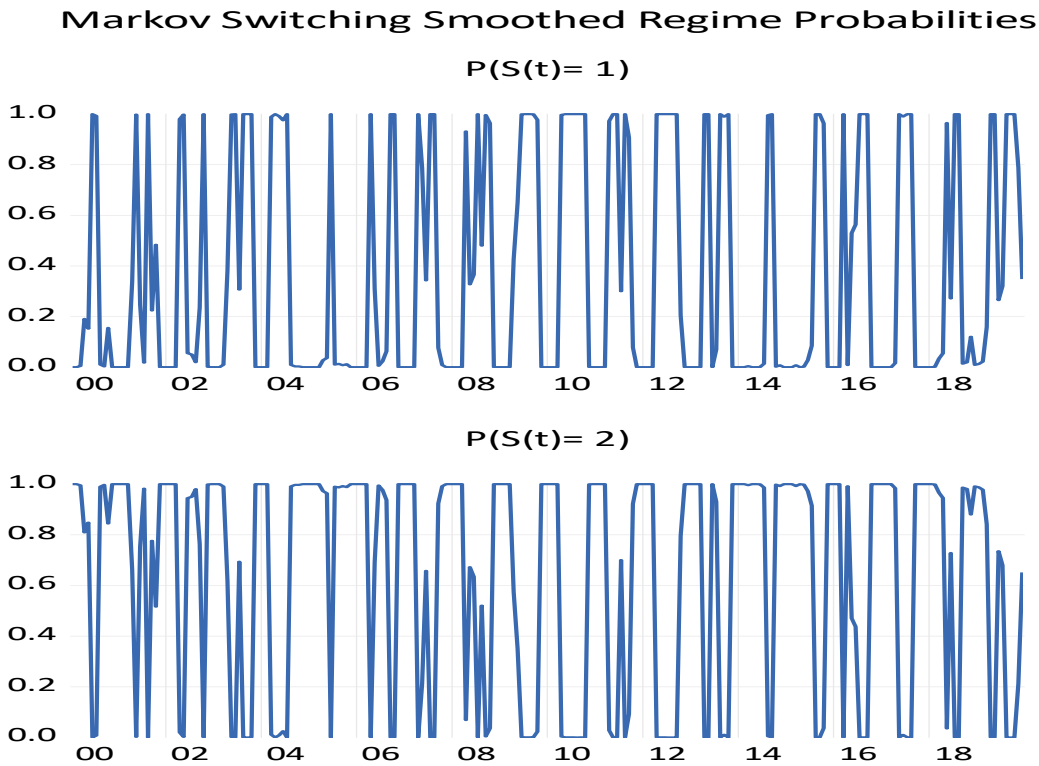

Figure 4.9: Smoothed Probabilities for the Navrongo Substation

In verifying the fit of the model, the residuals were analyzed. This is to ensure that the residuals are not autocorrected for any lag. The Durbin Watson statistical test was used to detect the presence of autocorrelation at lag 1. Durbin and Watson (1950, 1951) applied this test statistic to the residuals in analyzing the least square regressions, and developed bounds tests for the null hypothesis that the errors are serially uncorrelated against the alternative that they follow a first order AR process. The test statistic for the Navrongo weather substation is 2.000 showing that the residuals are uncorrelated. This is supported by the plot of the residuals. The residuals meet the properties of the residuals. The residuals basically form a "horizontal band" around the 0 line suggesting that the

variances of the error terms are constant. Likewise, no one residual peak from the stochastic pattern of the residuals, signifying that there are not outliers.

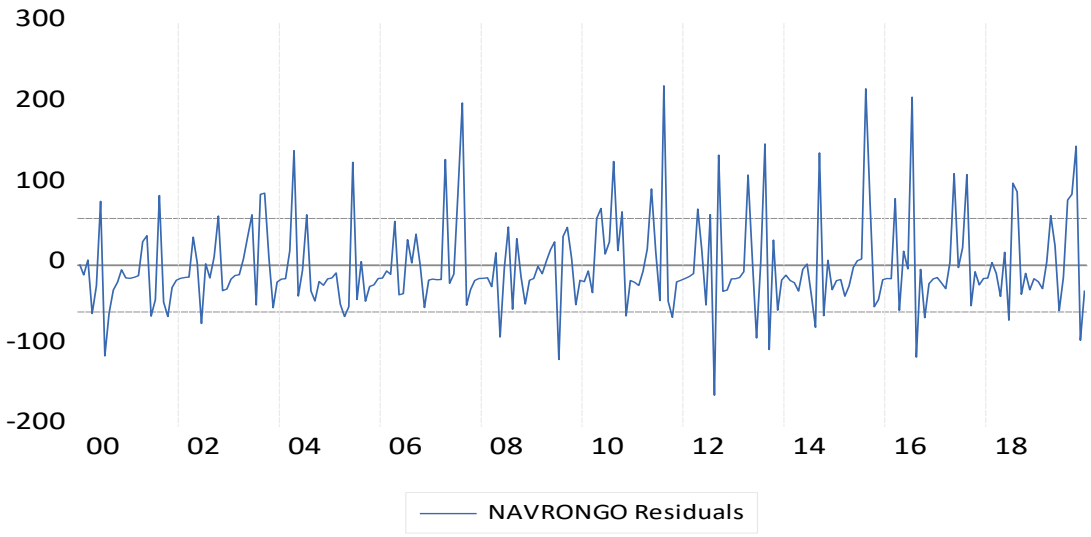

Figure 4.10: Residual plot of the Navrongo Substation

After the estimation of all the parameters of the MS-AR model, the model was revised to ensure its adequacy in representing the series. Figure 4.11 illustrated the actual series, the proposed fitted MS (2)-AR (1) model and the residuals. From the Figure, it can be observed that the proposed model is able to adequately capture the behavior of the index over time. The proposed model is able to identify the peaks and the drastic variations in the rainfall pattern in the study area. Likewise, the model is able to identify various regimes fluctuations associated with the rainfall in the substation. The time series plot of the proposed model follows the trends observed during the historical episodes. The model is able to reproduce the behavioral pattern of the phenomenon.

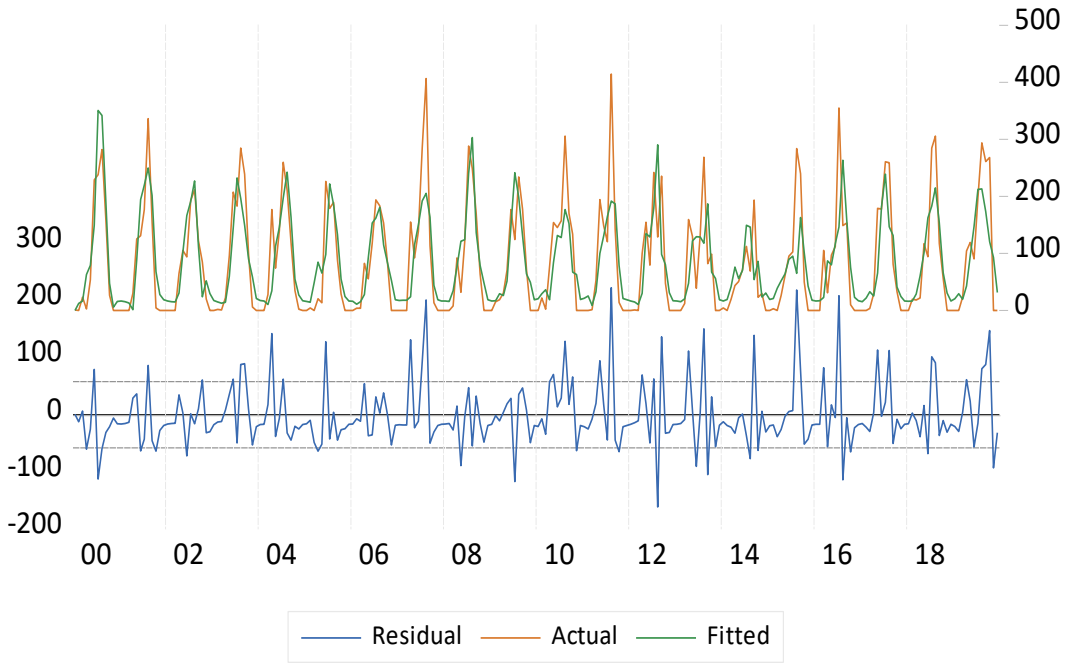

Figure 4.11: Actual, Fitted and Residual plot of the Navrongo Substation

4.5.2 Markov Switching Autoregressive Model for Bolgatanga Substation

The AR terms to include based on the PACF plot are lag 1 and lag 11. The MS (2)-AR (p) models at lag 1, and 11, are analyzed to determine the best model that represents the series. Table 4.11 displays the AIC, BIC and HQC values of models MS (2)-AR (1) and MS (2)-AR (11). The model MS (2)-AR (1) outperforms the model MS (2)-AR (11) in all the goodness of fit measures. The model MS (2)-AR (1) is therefore the preferred model.

Table 4.11: Model for Bolgatanga Substation

Model	AIC	BIC	HQC
MS (2) - AR (1)	10.73	10.83	10.77
MS (2) - AR (11)	11.31	11.39	11.34

Given the selected model MS (2)-AR (1), the estimated parameters are summarized in Table 4.12. The parameters μ_{s_t} which is a constant and depends on the state S_t are $\mu_1 = 3.00$ and $\mu_2 = 4.36$. Each one of these estimated coefficients is statistically significant at 5% level of significance. Also, the parameter δ_p is the coefficient of the autoregressive component which also depends on S_t. The coefficient of the AR (1) are $\delta_1 = 0.79$ and $\delta_2 = 0.26$ for regime 1 and 2 respectively. These coefficients are statistically significant at 5% level of significance.

Table 4.12: Estimates of Regime Parameters for Bolgatanga Substation

Regime	Parameter	Coefficient	Std. Error	z-Statistic	Prob.
1	μ	3.00	0.07	45.51	<0.0001
	δ	0.79	0.15	25.27	<0.0001
2	μ	4.36	0.12	35.16	<0.0001
	δ	0.26	0.01	32.52	<0.0001

The MSAR for the Navrongo weather substation can therefore be specified as;

$$y_t = \begin{cases} 0.79(y_{t-1} - 3.00), S_t = 1 \\ 0.26(y_{t-1} - 4.36), S_t = 2 \end{cases}.$$

The transitional probabilities, the filtered and smoothed probabilities are estimated. The transitional probabilities describe a stochastic process that undergoes transition from one state to the other. The transition probability p_{ij} illustrates the likelihood of transition to regime j after the occurrence of regime i. These transitional probabilities of the model MS (2)-AR (1) for series are $P_{11} = 0.82$, $P_{12} = 0.18$, $P_{21} = 0.27$, and $P_{22} = 0.73$. These transitional probabilities are organized and represented in a transitional probability matrix as;

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 0.82 & 0.18 \\ 0.27 & 0.73 \end{pmatrix}.$$

The transitional probabilities estimate the probabilities of moving from one state to the other within the state space. The transitional probabilities satisfy the condition $p_{i1} + p_{i2} + ... + p_{iM} = 1$. From the transitional probabilities, P_{11} represents the probability of experiencing rainfall for a known period of high rainfall. That is, during a known period of high rainfall, the probability of experiencing rainfall is estimated to be approximately 82% within the Bolgatanga enclave. Also, during a period of high rainfall, the probability of experiencing little or no rainfall is estimated to be 18%. Furthermore, the probability of experiencing high rainfall given that the previous state experienced low or no rainfall is 27%. The probability of experiencing no or low rainfall given the low

or no rainfall experienced previously is estimated to be 73%. The plots of these transitional probabilities are displayed in Figure 4.12.

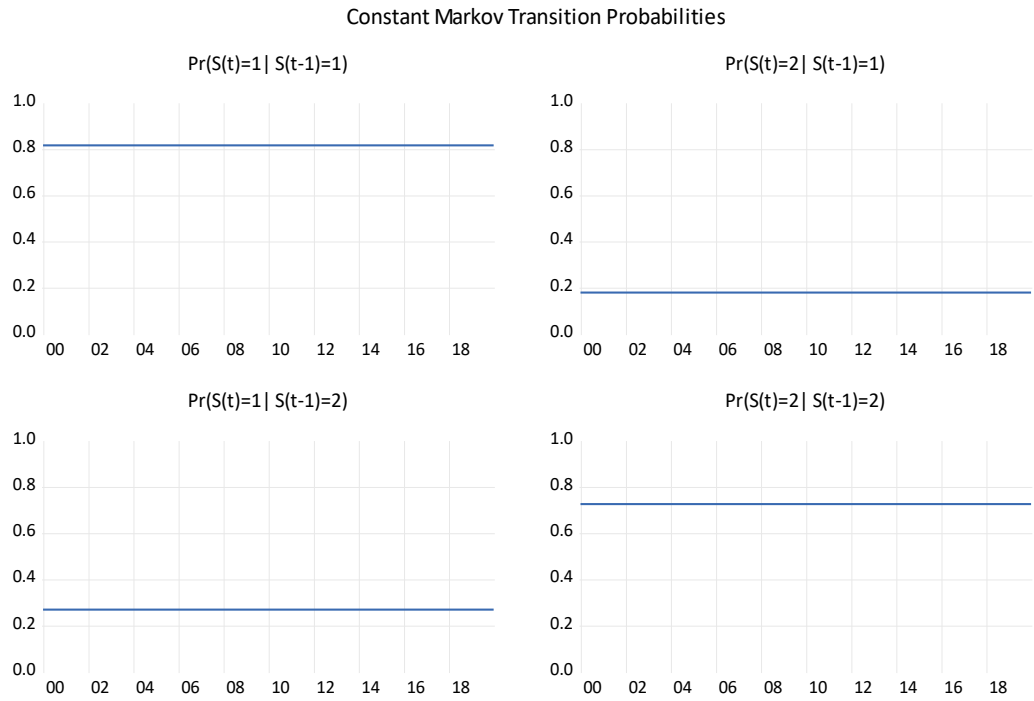

Figure 4.12: Constant Markov Transitional Probabilities for Bolgatanga Substation

The expected duration estimates how long it takes for each phenomenon to stay in each regime before transiting into another regime. It is estimated that the expected duration for rainfall in the study area is 5.5 months and that for low or no rainfall is estimated to be at least 3.7 months. This is consistent with the historical data gathered about the study area. The Bolgatanga weather substation records rainfall in substantial amount from the month of April and reaches its peak by the month of August. This goes to confirm observations from table 4.2. The

rainfall declines in magnitude afterwards towards November. The plots of these expected durations are displayed in Figure 4.13.

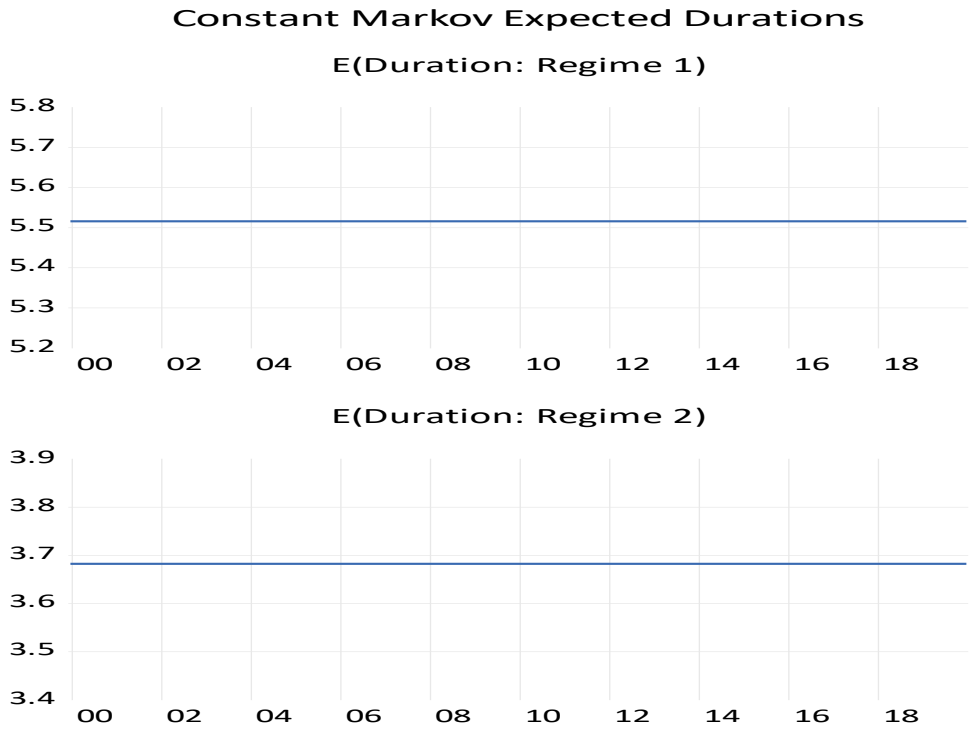

Figure 4.13: Constant Markov Expected Duration for Bolgatanga Substation

The smoothing probabilities $IP(s_t = i \mid Z^T; \theta)$, is computed using the method of Kim (1994). The estimation values are obtained based upon the computation of the smoothened probabilities, employing all the observational data until $t = T$. The probability value of the regimes (S_t) with value j based on observations until time $t = T$ is also computed. Through the process of filtering and smoothing the

probability of y_t at every value of the state s_t is obtained. The plots of these filtered and smoothed probabilities are as in Figure 4.14.

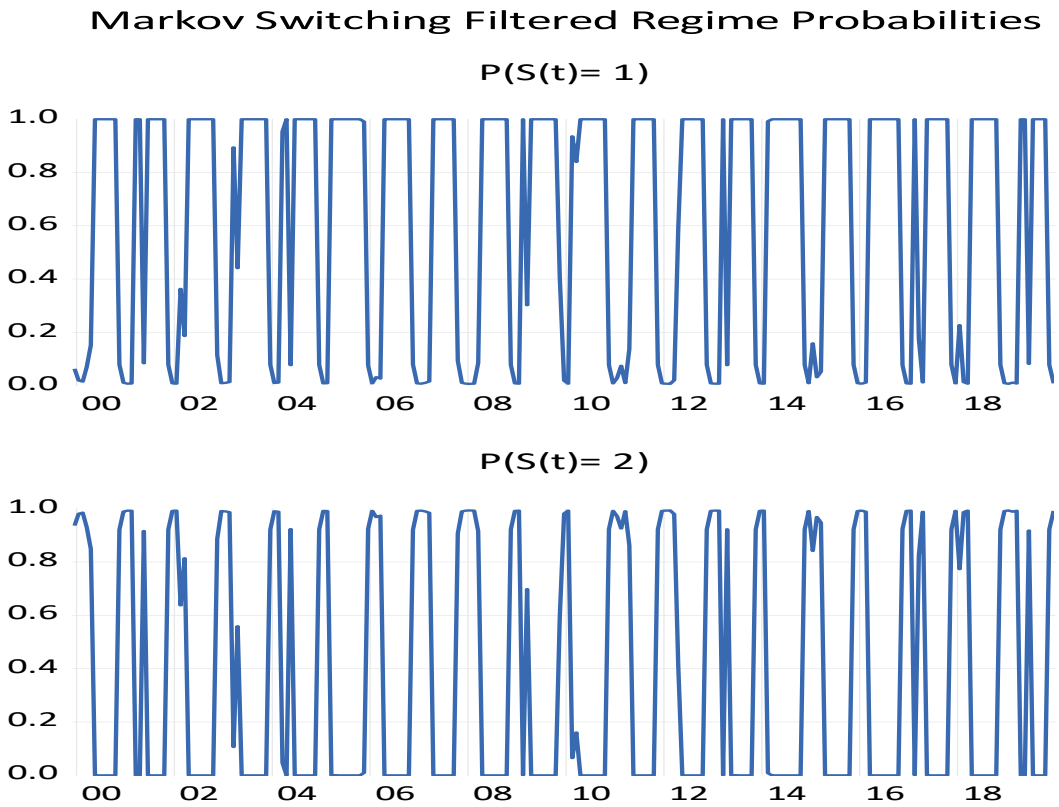

Figure 4.14: Filtered Probabilities for the Bolgatanga Substation

The transition probabilities of remaining in the same state $P(S(t)=1), P(S(t)=2)$ are approximated to one or zero for various seasons. Each of these periods or seasons represents a cycle of the rainy season and the dry season. During the rainy season, the probability of remaining in regime one $P(S(t)=1)$ is close to one. This is consistent with the estimates from the transitional probabilities. Thus given a period of high rainfall, the chances of experiencing high rainfall in a month is estimated to be at $P_{11}=0.82$. Conversely, during the dry season or period of minimal rainfall, the probability of remaining in regime 2 thus $P(S(t)=2)$ is

close to zero. This thus suggests that these regimes are mutually exclusive. As it can be expected, the same pattern can be observed throughout the period of the study. The plot of the smoothed probabilities for the Navrongo weather substation is in Figure 4.15.

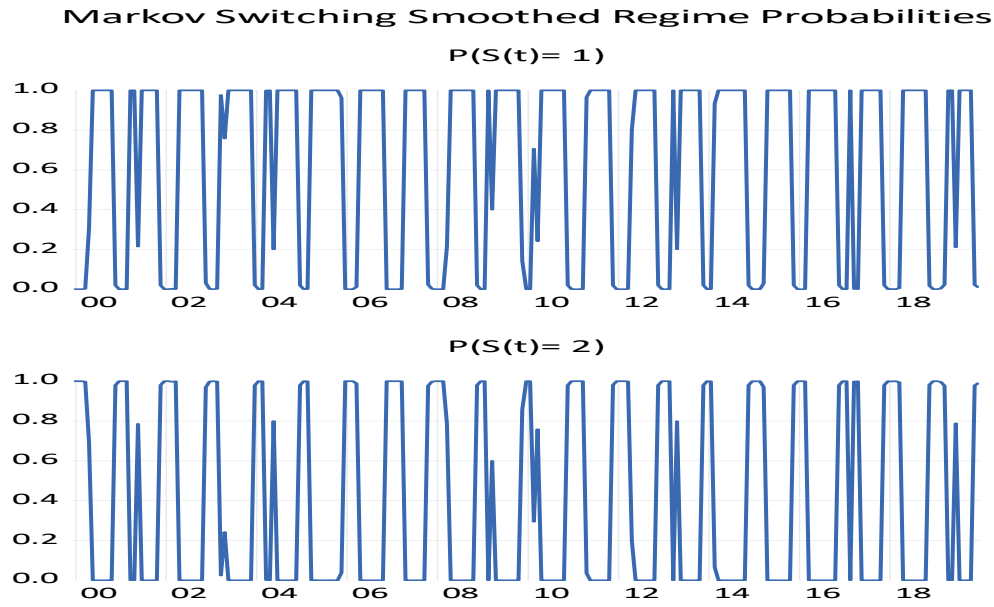

Figure 4.15: Smoothed Probabilities for the Bolgatanga Substation

The residual are examined to ensure that they are not autocorrelated. The Durbin Watson test statistic is used to diagnose the autocorrelation. The Durbin Watson statistic refers to a test statistic used to detect the presence of autocorrelation at lag 1 in the residuals. The test statistic for the Bolgatanga weather substation is 2.003 showing that the residuals are uncorrelated. The plots of the residuals are displayed in Figure 4.16. The fitted model and the residual plot is in Figure 4.17.

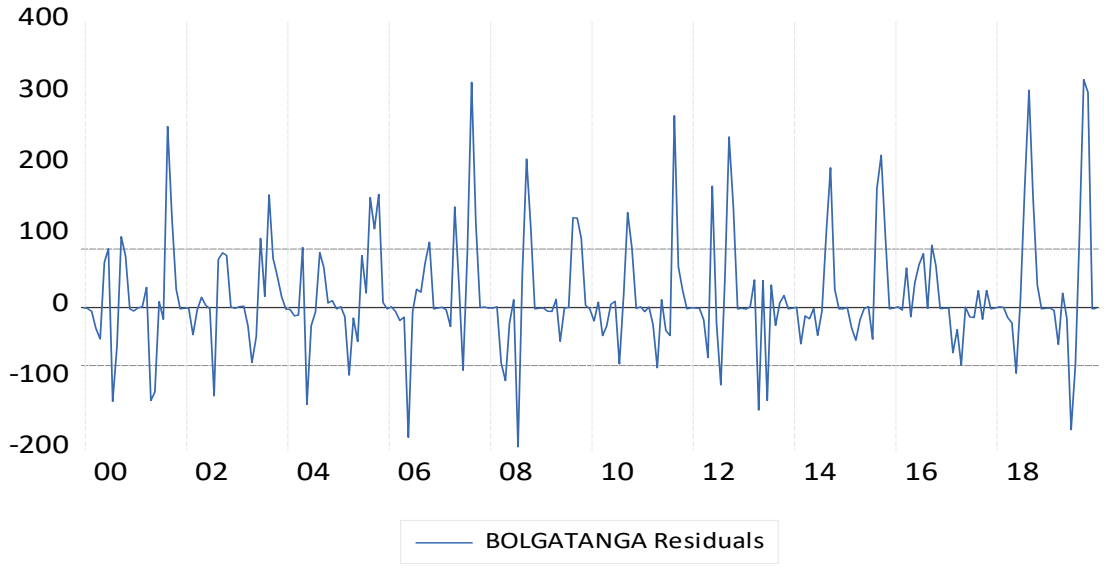

Figure 4.16: Residual plot of the Bolgatanga Substation

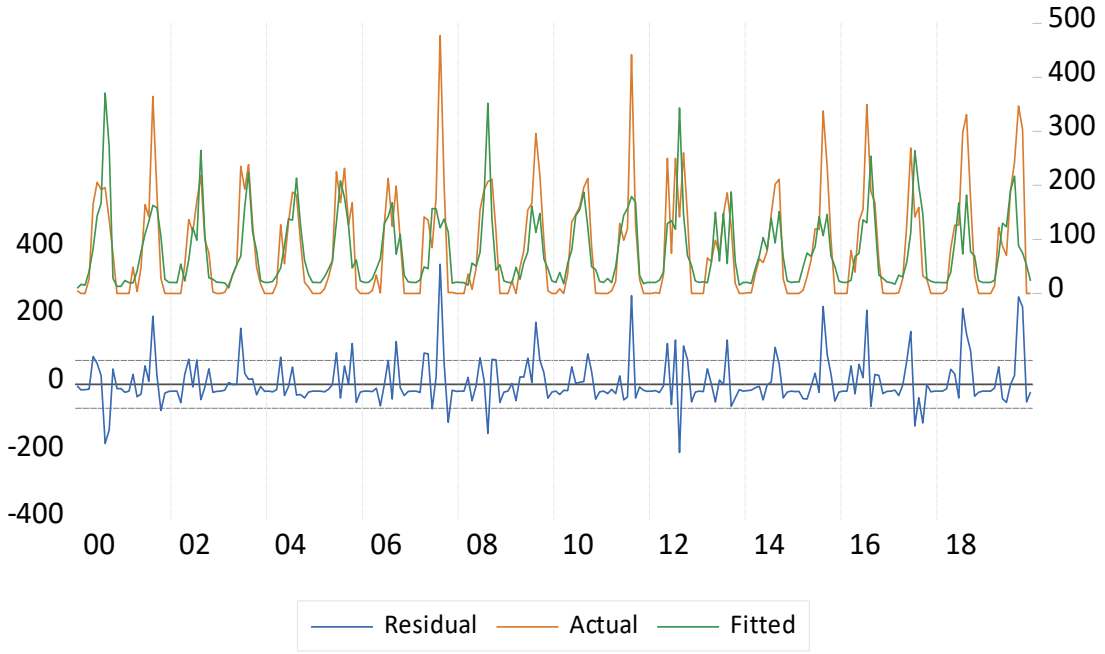

Figure 4.17: Actual, Fitted and Residual plot of the Bolgatanga Substation

4.6 Discussions of Results

The study findings revealed that the expected duration for rainy and dry season are 5.1 and 3.05 months respectively for the Navrongo substation and 5.5 and 3.7 months for the Bolgatanga substation respectively. In between these extreme regimes are the transitional periods. These are transition from rainy season to dry season and transition from dry season to rainy season. The expected durations are at least 1.6 and 2.1 months for transition from rainy to dry season and from dry season to rainy season respectively for the Navrongo substation. The transitional periods for the Bolgatanga weather substation are 1.2 and 1.4 months for transition from rainy to dry season and transition from dry to rainy season respectively. The rainfall behavior as exhibited in the study findings showed that the area receives a single rainfall maxima and for an average period of about 5 months. These months do not receive equal amount and an even distribution of rainfall as shown in figure 4.1. Consistent with Asamoah et al. (2020), parts of the months classified as rainy season receives rainfall quantities that are not sufficient to support crop and plants growth. It is evident that in the study area rainfall is averagely low, with erratic pattern. It also fluctuates and varies in time with some extremely low and high periods of rainfall as seen in figure 4.1. This could be as a result of the consequences of climate change which causes variability in weather events. This is supported by Owusu and Waylen (2013), who concludes that one of the effects of climate change is the fluctuation of rainfall with extreme rainfall such as erratic rainfall and high rainfall. Also, Ampadu et al. (2019) in their study about the distribution of rainfall in the Upper East Region concluded that even

though rainfall occurs from May to September, the month of July, August, and September records high rainfall with low variation. This explains that even though rainfall spans the period May to September, months that experience enough rainfall to support plant and crop growth are July, August, and September since there is reliable occurrence of precipitation within this period of the year.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.0 Introduction

This chapter presents the summary, conclusions and recommendations of the study. The chapter covers the summary of the findings from the data analyzed, conclusions and recommendations based on the findings from the data.

5.1 Summary

The study investigated the pattern of rainfall distribution in the Navrongo and Bolgatanga Municipal using secondary data obtained from Ghana Meteorological Agency from January 1999 to December 2019. The study also aimed at modeling the rainfall in the study locations.

The results of the preliminary analyses indicated that the minimum rainfall recorded for all the stations was 0.00 mm. The maximum rainfall recorded was 455.50 mm and 476.90 for Navrongo and Bolgatanga weather substations respectively. The mean rainfall being 84.39 mm and 83.97 mm for Navrongo and Bolgatanga weather substations respectively. The skewness measures are above 0, indicating positively skewed data sets. The measure of excess kurtosis for the Navrongo and Bolgatanga weather substations is 0.73 and 1.26 respectively.

The concentration of rainfall is along the months of April, May, June, July, August, September and October while the month of January, February, March, November and December experiences low or no rainfall.

The stationary behavior of the data was explored. The data sets were found to be non-stationary. However, both data sets were stationary after differencing of the data sets once.

To successfully model the Navrongo and Bolgatanga rainfall using MS-AR models, obtaining a lag order which will give an optimal autoregressive model for modeling the data is crucial. The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots were used to identify the number of the autoregressive terms p that are significant and to be included in the estimation of the parameters. That is the autoregressive coefficients and the transitional probabilities. State one was classified as rainy season while state two was classified as dry season.

For further analysis of the data sets, MS-AR models were used. MS (2)-AR (1) was obtained as the best model for the Navrongo substation as it had the least of goodness-of-fit measures. Also for the Bolgatanga data sets, MS (2)-AR (1) was obtained as the best model.

The transitional probabilities for the Navrongo weather substation are $P_{11} = 0.80$, $P_{12} = 0.20$, $P_{21} = 0.35$, and $P_{22} = 0.65$. This explains that given a period of rainfall, the chances of experiencing rainfall over a month will be high at $P_{11} = 0.80$, and the likelihood of rainfall declines to $P_{12} = 0.20$ for periods of no rainfall. Moreover, for a period of no rainfall, the likelihood of rainfall experienced over the month is $P_{22} = 0.65$. The probability of experiencing rainfall

in a particular period is estimated at $P_{21} = 0.35$ given that the preceding period experienced no or minimal rainfall. It is estimated that when the phenomenon is in regime one, it is expected to spend at least 5.1 months in that regime. A switch then happens and the phenomenon is expected to stay in that regime for at least 3.05 months before another switch takes place. The smoothed and filtered probabilities indicated that the value of the probability is close to one when a phenomena is being experienced and is close to close when the phenomena is not experienced.

The transitional probabilities for the Bolgatanga weather substation are $P_{11} = 0.82$, $P_{12} = 0.18$, $P_{21} = 0.27$, and $P_{22} = 0.73$. This explains that during a known period of high rainfall, the probability of experiencing rainfall is estimated to be approximately 81.87% within the Bolgatanga enclave. Also, during a period of high rainfall, the probability of experiencing no rainfall is estimated to be 18.13%. Furthermore, the probability of experiencing high rainfall given that the previous state experienced low or no rainfall is 27.16%. The probability of experiencing no or low rainfall given the low or no rainfall experienced previously is estimated to be 72.84%. It is estimated that the expected duration for rainfall in the study area is 5.5 months and that for dry season is estimated to be at least 3.7 months.

5.2 Conclusions

The study shows that a two state MS-AR model best describes the pattern of rainfall in the weather substations in the Upper East Regions. MS (2)-AR (1) and MS (2)-AR (1) were the best models for modeling the data sets from Navrongo and Bolgatanga. The study further reveals that during a rainfall regime, the probability of remaining in the regime is estimated to be 80% for the Navrongo and 82% for the Bolgatanga substation. The expected duration for a rainfall regime is 5.1 months and 5.5 months for the Navrongo and Bolgatanga substations respectively. Also, the probability of rainfall is reduced given the previous month received low rainfall. The expected duration of dry season is 3.1 and 3.7 months for the Navrongo and Bolgatanga substations respectively.

5.3 Recommendations

The study makes the following recommendations based on the findings.

i. The study result shows a variation in the expected duration for the various regimes. It is recommended that government assists farmers with early maturing seedlings and drought resistant seedlings. This will enable farmers cultivate and have good harvest without a substantial impact of the climate variability and climate change.

ii. The study estimated the expected duration for each regime. Government through its agriculture extension officers should educate farmers on the right time to plant as is often the most important part of the farming. Planting at the right time will result in higher yields as farmers will be able to take advantage of the favorable farming conditions within the rainfall regime.

iii. The study findings show that the amount of rainfall received is not evenly distributed during the rainy season. Farmers should therefore ensure good water management in their farms as it is essential to crop survival and the maximization of crop yield potential. That is farmers should put in mechanism to ensure the crops are getting enough water, but also that they are not being over-watered. Proper drainage systems within the farms should be developed to help prevent waterlogging and salinization in the soil which can stifle growth and production.

REFERENCES

Abdul-Aziz, A. R., Anokye, M., Kwame, A., Munyakazi, L., and Nsowah-Nuamah, N. N. N. (2013). Modeling and forecasting rainfall pattern in Ghana as a seasonal ARIMA process: The case of Ashanti region. *International Journal of Humanities and Social Science*, *3*(3), 224-233.

Ahunu, L., and Ackah, I. (2017). Towards Ghana's Energy Security: An Analyses of Existing Natural Gas Legal Frameworks in Ghana. *Available at SSRN 2928266*.

Alagidede, P., Baah-Boateng, W., and Nketiah-Amponsah, E. (2013). The Ghanaian economy: an overview. *Ghanaian Journal of Economics*, *1*(1), 4-34.

Alvarez, R., Camacho, M., and Ruiz, M. (2019). Inference on filtered and smoothed probabilities in Markov-switching autoregressive models. *Journal of Business & Economic Statistics*, *37*(3), 484-495.

Ampadu, B., Sackey, I., and Cudjoe, E. (2019). Rainfall Distribution in the Upper East Region of Ghana, 1976–2016. *Ghana Journal of Science, Technology and Development*, *6*(2), 45-59.

Ang, A., and Timmermann, A. (2012). Regime changes and financial markets. *Annu. Rev. Financ. Econ.*, *4*(1), 313-337.

Aninagyei, I., and Appiah, D. O. (2014). Analysis of Rainfall and Temperature Effects on Maize and Rice Production Akim Achiase, Ghana.

Asamoah, Y., and Ansah-Mensah, K. (2020). Temporal Description of Annual Temperature and Rainfall in the Bawku Area of Ghana. *Advances in Meteorology*, *2020*.

Auffhammer, M., Hsiang, S. M., Schlenker, W., and Sobel, A. (2011). Global climate models and climate data: a user guide for economists. *Unpublished manuscript*, *1*, 10529-10530.

Awojobi, O. N., and Tetteh, J. (2017). The impacts of climate change in Africa: a review of the scientific literature. *Journal of International Academic Research for Multidisciplinary*, *5*(11), 39-52.

Ayodeji, I. O. (2016). A three-state Markov-modulated switching model for exchange rates. *Journal of Applied Mathematics*, *2016*.

Bazzi, M., Blasques, F., Koopman, S. J., and Lucas, A. (2017). Time-varying transition probabilities for Markov regime switching models. *Journal of Time Series Analysis*, *38*(3), 458-478.

Berhane, T., Shibabaw, N., Awgichew, G., and Kebede, T. (2018). Modeling and Forecasting Rainfall in Ethiopia. *International Journal of Computing Science and Applied Mathematics*, *4*(2), 42-46.

Biao, E. I. (2017). Assessing the impacts of climate change on river discharge dynamics in Oueme River Basin (Benin, West Africa). *Hydrology*, *4*(4), 47.

Boot, T., and Pick, A. (2014). Optimal forecasts from Markov switching models and the effect of uncertain break dates.

Bordon, A. R., and Weber, A. (2010). The transmission mechanism in Armenia: New evidence from a regime switching VAR analysis. *IMF Working Papers*, 1-31.

Buadi, D. K., Anaman, K. A., and Kwarteng, J. A. (2013). Farmers' perceptions of the quality of extension services provided by non-governmental organisations in two municipalities in the Central Region of Ghana. *Agricultural Systems*, *120*, 20-26.

Bukari, D., Tuokuu, F. X. D., Suleman, S., Ackah, I., and Apenu, G. (2020). Ghana's energy access journey so far: a review of key strategies. *International Journal of Energy Sector Management*.

Camacho, M., Perez-Quiros, G., and Poncela, P. (2012). Markov-switching dynamic factor models in real time.

Cárdenas-Gallo, I., Sánchez-Silva, M., Akhavan-Tabatabaei, R., and Bastidas-Arteaga, E. (2015, July). A Markov Regime-Switching Framework Application for Describing El Niño Southern Oscillation (ENSO) Patterns.

Chan, K. S., Hansen, B. E., & Timmermann, A. (2017). Guest editors' introduction: Regime switching and threshold models.

Chang, Y., Choi, Y., and Park, J. Y. (2014). Regime Switching Model with Endogenous Autoregressive Latent Factor. *Manuscript, Indiana University*.

Chang, Y., Choi, Y., and Park, J. Y. (2017). A new approach to model regime switching. *Journal of Econometrics*, *196*(1), 127-143.

Cheung, Y. W., and Lai, K. S. (1995). Lag order and critical values of the augmented Dickey–Fuller test. *Journal of Business & Economic Statistics*, *13*(3), 277-280.

Chowdhury, A. K., Lockart, N., Willgoose, G., Kuczera, G., Kiem, A. S., and Manage, N. P. (2017). Development and evaluation of a stochastic daily rainfall model with long-term variability. *Hydrology and Earth System Sciences*, *21*(12), 6541.

De Bon, H., Parrot, L., and Moustier, P. (2010). Sustainable urban agriculture in developing countries. A review. *Agronomy for sustainable development*, *30*(1), 21-32.

Devianto, D., Wisza, U. A., Wara, M., Permathasari, P., and Zen, R. O. M. (2018, September). Time Series of Rainfall Model with Markov Switching Autoregressive. In *2018 International Conference on Applied Information Technology and Innovation (ICAITI)* (pp. 202-207). IEEE.

Duprey, T., and Klaus, B. (2017). How to predict financial stress? An assessment of Markov switching models.

Durbin, J., and Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. *Biometrika*, *37*(3/4), 409-428.

Durbin, J., and Watson, G. S. (1971). Testing for serial correlation in least squares regression. III. *Biometrika*, *58*(1), 1-19.

Emmanuel, L. A., Hounguè, N. R., Biaou, C. A., and Badou, D. F. (2019). Statistical Analysis of Recent and Future Rainfall and Temperature Variability in the Mono River Watershed (Benin, Togo). *Climate*, *7*(1), 8.

Evarest, E., Berntsson, F., Singull, M., and Charles, W. (2016). Regime switching models on temperature dynamics.

Exterkate, P., and Knapik, O. (2017). *A regime-switching stochastic volatility model for forecasting electricity prices* (No. 2017-03). Department of Economics and Business Economics, Aarhus University.

Filardo, A. J. (1998). *Choosing information variables for transition probabilities in a time-varying transition probability Markov switching model* (No. 98-09). Federal Reserve Bank of Kansas City.

Filder, T. N., Muraya, M. M., and Mutwiri, R. M. (2019). Application of seasonal autoregressive moving average models to analysis and forecasting of time series monthly rainfall patterns in Embu County, Kenya. *Asian Journal of Probability and Statistics*, 1-15.

Friesen, J., and Diekkriuger, B. (2002). Spatio-temporal rainfall patterns in Northern Ghana. *Diplom Thesis, Geographische Institute der Rheinischen Friedrich-Wilhelms, University at Bonn, Bonn*.

Ghana. Statistical Service. (2011). *Multiple Indicator Cluster Survey, 2011: Monitoring the Situation of Children, Women, and Men; with an Enhanced Malaria Module and Biomarker*. Ghana Statistical Service.

Gujarati, D. N., and Porter, D. C. (1999). *Essentials of econometrics* (Vol. 2). Singapore: Irwin/McGraw-Hill.

Gujarati, D. N., and Porter, D. C. (2003). Basic econometrics (ed.). *Singapore: McGrew Hill Book Co*.

Gyamerah, S. A., Ngare, P., and Ikpe, D. (2018). Regime-switching temperature dynamics model for weather derivatives. *International Journal of Stochastic Analysis*, *2018*.

Gyamfi, E. A. (2019). Report of the Committee on Mines and Energy on the 2020 annual budget estimates of the Ministry of Energy.

Hayashi, F., and Koeda, J. (2013). A regime-switching SVAR analysis of quantitative easing. *CARF F-Series CARF-F-322, Center for Advanced Research in Finance, Faculty of Economics, The University of Tokyo*, *12*.

Holzkämper, A., Calanca, P., and Fuhrer, J. (2012). Statistical crop models: predicting the effects of temperature and precipitation changes. *Climate Research*, *51*(1), 11-21.

Issahaku, A. R., Campion, B. B., and Edziyie, R. (2016). Rainfall and temperature changes and variability in the Upper East Region of Ghana. *Earth and Space Science*, *3*(8), 284-294.

Joanes, D. N., and Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*(1), 183-189.

Kaufmann, S. (2016). Hidden Markov models in time series, with applications in economics.

Kole, E. (2019). Markov Switching Models: An Example for a Stock Market Index. *Available at SSRN 3398954*.

Kole, E. (2019). Markov Switching Models: An Example for a Stock Market Index. *Available at SSRN 3398954*.

Kuan, C. M. (2002). Lecture on the Markov switching model. *Institute of Economics Academia Sinica*, 1-30.

Kumi, E. N. (2017). *The electricity situation in Ghana: Challenges and opportunities*. Washington, DC: Center for Global Development.

Kyei-Mensah, C., Kyerematen, R., and Adu-Acheampong, S. (2019). Impact of Rainfall Variability on Crop Production within the Worobong Ecological Area of Fanteakwa District, Ghana. *Advances in Agriculture*, *2019*.

Mendy, D., and Widodo, T. (2018). Two Stage Markov Switching Model: Identifying the Indonesian Rupiah Per US Dollar Turning Points Post 1997 Financial Crisis.

Mollicone, D., Freibauer, A., Schulze, E. D., Braatz, S., Grassi, G., and Federici, S. (2007). Elements for the expected mechanisms on 'reduced emissions from deforestation and degradation, REDD'under UNFCCC. *Environmental Research Letters*, *2*(4), 045024.

Nkrumah, F., Klutse, N. A. B., Adukpo, D. C., Owusu, K., Quagraine, K. A., Owusu, A., and Gutowski, W. (2014). Rainfall variability over Ghana: model versus rain gauge observation. *International Journal of Geosciences*, *5*(7), 673.

Nyatuame, M., and Agodzo, S. K. (2018). Stochastic ARIMA model for annual rainfall and maximum temperature forecasting over Tordzie watershed in Ghana. *Journal of Water and Land Development*, *37*(1), 127-140.

Olatayo, T. O., and Taiwo, A. I. (2014). Statistical modelling and prediction of rainfall time series data. *Global Journal of Computer Science and Technology*.

Oseifuah, E. K., and Korkpoe, C. H. (2019). A Markov regime switching approach to estimating the volatility of Johannesburg Stock Exchange (JSE) returns. *Investment management and financial innovations*, (16, Iss. 1), 215-225.

Oteng-Abayie, E. F., and Dramani, J. B. (2019). Time-frequency domain causality of prime building cost and macroeconomic indicators in Ghana: implications for project selection. *Construction Management and Economics*, *37*(5), 243-256.

Owusu, K., and Klutse, N. A. B. (2013). Simulation of the Rainfall Regime over Ghana from CORDEX.

Owusu, K., and Waylen, P. R. (2013). Identification of historic shifts in daily rainfall regime, Wenchi, Ghana. *Climatic Change*, *117*(1-2), 133-147.

Pendergrass, A. G., Knutti, R., Lehner, F., Deser, C., and Sanderson, B. M. (2017). Precipitation variability increases in a warmer climate. *Scientific reports*, *7*(1), 1-9.

Perron, P. (1988). Trends and random walks in macroeconomic time series: Further evidence from a new approach. *Journal of economic dynamics and control*, *12*(2-3), 297-332.

Piger, J. (2009). Econometrics: Models of Regime Changes.

Rhyne, A. L., Tlusty, M. F., Schofield, P. J., Kaufman, L. E. S., Morris Jr, J. A., and Bruckner, A. W. (2012). Revealing the appetite of the marine aquarium fish trade: the volume and biodiversity of fish imported into the United States. *PLoS One*, *7*(5), e35808.

Salhi, K., Deaconu, M., Lejay, A., Champagnat, N., and Navet, N. (2016). Regime switching model for financial data: Empirical risk analysis. *Physica A: Statistical Mechanics and its Applications*, *461*, 148-157.

Shin, Y., and Schmidt, P. (1992). The KPSS stationarity test as a unit root test. *Economics Letters*, *38*(4), 387-392.

Smith, A., Naik, P. A., and Tsai, C. L. (2006). Markov-switching model selection using Kullback–Leibler divergence. *Journal of Econometrics*, *134*(2), 553-577.

Song, Y., and Woźniak, T. (2020). Markov Switching. *arXiv preprint arXiv:2002.03598*.

Torgbor, F. F., Stern, D. A., Nkansah, B. K., and Stern, R. D. (2018). Rainfall Modelling with a Transect View in Ghana. *Ghana Journal of Science*, *58*, 41-57.

Van Norden, S., and Vigfusson, R. J. (1996). Regime-switching models: A guide to the Bank of Canada Gauss procedures.

Watson, G. S., and Durbin, J. (1951). Exact tests of serial correlation using noncircular statistics. *The Annals of Mathematical Statistics*, 446-451.

Wuebbles, D. J., Fahey, D. W., and Hibbard, K. A. (2017). Climate science special report: fourth national climate assessment, volume I.