UNIVERSITY OF ENERGY AND NATURAL RESOURCES, SUNYANI

THE VARIABILITY CLIMATE CHANGE IS RESPONSIBLE FOR IN VEGETATION LOSS IN GHANA

KALONG BONIFACE

FUGAH SELETEY MITCHELL

DEPARTMENT OF MATHEMATICS AND STATISTIC

SCHOOL OF SCIENCE

JUNE 2019

THE VARIABILITY CLIMATE CHANGE IS RESPONSIBLE FOR IN VEGETATION LOSS IN GHANA

by

Name - exams number

B.Sc. (Program) Location

A Thesis submitted to the Department of Mathematics and Statistics, School of Science, University of Energy and Natural Resources, Sunyani in partial fulfillment of the requirements for the degree of Bachelor of Science in Mathematics

JUNE 2019

DECLARATION AND CERTIFICATION

Certified that the thesis entitled "**THE VARIABILITY CLIMATE CHANGE IS RESPONSIBLE FOR IN VEGETATION LOSS IN GHANA**", submitted by **Kalong Boniface** and **Fugah Seletey Mitchell** to the **DEPARTMENT** , for the award of the degree of Doctor of Philosophy has been accepted by the external examiners and that the student has successfully defended the thesis in the viva voce examination held today.

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

Supervisor's Certification

This study was carried out under the supervisory committee of (Names of all Supervisors) in accordance with the guidelines on supervisions of graduate studies.

Major Supervisor's Name and Qualifications

Signature: ..... Date: .....

Co-Supervisor's Certification

Co-Supervisor's Name and Qualifications

Signature: ..... Date: .....

Co-Supervisor's Name and Qualifications

Signature: ..... Date: .....

ABSTRACT

All thing change, but how we respond to change is our responsibility, to fare it or embrasse it. Resisting change leads to one fiat. Our own extinction. Time is a smybole of freedom and peace

DEDICATION

Write dedication here

ACKNOWLEDGMENTS

Write Acknowledgment here

Contents

List of Tables

List of Figures

CHAPTER ONE

1.0 INTRODUCTION

1.1 Introduction

One would anticipate that the majority of emerging nations, which are still in the early stages of economic development and growth, would have a high forest cover and little deforestation. This, however, has not been the case. Ghana is a lower-middle-income nation that is still working toward middle-income classification. However, it has already begun to see a deforestation rate that is comparable to that of middle-income countries. The rapid population expansion, clearing of field for Galamsey operation,increased domestic need of wood for things like fuel, furniture, construction, and timber exports have all contributed to this trend, Bush fires in the 1980s, climate change, and lax law enforcement have all had an impact.

The purpose of this paper is to establish an understanding in time series analysis on remotely sensed data. Which will introduced us to the fundamentals of time series modeling, including decomposition, autocorrelation and modeling historical changes in Galamsey Operation in Ghana, the Cause,Dangers and it's Environmental impact.

Galamsey also known as "gather them and sell",(**Owusu-Nimo2018**) is the term given by local Ghanaian for illegal small-scale gold mining in Ghana . The major cause of Galamsey is unemployment among the youth in Ghana (**Gracia2018**). Young university graduates rarely find work and when they do it hardly sustains them. The result is that these youth go the extra mile to earn a living for themselves and their family.

Another factor is that lack of job security. On November 13, 2009 a collapse occurred in an illegal, privately owned mine in Dompoase, in the Ashanti Region of Ghana. At least 18 workers were killed, including 13 women, who worked as porters for the miners. Officials described the disaster as the worst mine collapse in Ghanaian history (**womendi2009**).

Illegal mining causes damage to the land and water supply (**Ansah2017**) . In March 2017, the Minister of

Lands and Natural Resources, Mr. John Peter Amewu, gave the Galamsey operators/illegal miners a three-week ultimatum to stop their activities or be prepared to face the law (**Allotey2017**) . The activities by Galamseyers have depleted Ghana's forest cover and they have caused water pollution, due to the crude and unregulated nature of the mining process (**gyekye**).

Under current Ghanaian constitution, it is illegal to operate as galamseyer.That is to dig on land granted to mining companies as concessions or licenses and any other land in search for gold. In some cases, Galamseyers are the first to discover and work extensive gold deposits before mining companies find out and take over. Galamseyers are the main indicator of the presence of gold in free metallic dust form or they process oxide or sulfide gold ore using liquid mercury.

Between 20,000 to 50,000, including thousands from China are believed to be engaged in Galamsey in Ghana.But according to the Information Minister 200,000 and nearly 3 million people, recently are now into Galamsey operation and rely on it for their livelihoods (**goldgu2017**). Their operations are mostly in the southern part of Ghana where it is believe to have substantial reserves of gold deposits, usually within the area of large mining companies (**Barenblitt2021**) . As a group, they are economically disadvantaged. Galamsey settlements are usually poorer than neighboring agricultural villages. They have high rates of accidents and are exposed to mercury poisoning from their crude processing methods. Many women are among the workers, acting mostly as porters for the miners.

1.2 Background of The Study

As Galamsey is considered an illegal activity, they operations are hidden to the eyes of the authorities.So locating them is quite tricky ,but with satellite imagery ,it now possible to locate their operating and put an end to it. One of the features of Google Earth Engine is the ability to access years of satellite imagery without needing to download, organize, store and process this information. For instance, within the Satellite image

collection, now it possible to access imagery back to the 90's, allowing us to look at areas of interest on the map to visualize and quantify how much things has changed over time. With Earth Engine, Google maintains the data and offers it's computing power for processing.Users can now access hundreds of time series images and analyze changes across decades using GIS and R or other programming language to analyze these datasets.

1.3 Problem Statement

The Footprint of Galamsey is Spreading at a very faster rate, causing vegetation loss.Other factors accounting to vegetation loss may largely include climate change,urban and exurban development, bush fires. But not much works or research has been done to tell the extent to which Galamsey causes vegetation loss. This research attempts to segregate the variability climate is responsible for in vegetation loss so as to attribute the residual variability to Galamsey and other related activities such as bush-fires etc.

1.4 Research Questions

To address the challenge of the vegetation variability in this work, the following several statements were formed:

- Are there any changes in vegetation cause by Galamsey and Climate change in Ghana?
- Is there any relationship between vegetation and land surface temperature in Ghana?

1.5 Research Objectives

The purpose is to establish an understanding in time series analysis on remotely sensed data. We will be introduced to the fundamentals of time series modeling, including decomposition, auto-correlation and modeling historical changes.

- Perform time series analysis on satellite derived vegetation indices
- Estimate the extent to which Galamsey causes vegetation loss in Ghana.
- Dissociate or single out the variability climate is responsible for in vegetation loss

1.6 Significance Of The Study

There have been significant changes in vegetation cover in Ghana over the past 30 years, and these dynamics are related strongly to climatic factors such as temperature and other factors. In this study, we want to examine the effects of climatic change on Ghana's vegetation during these thirty years.
This study allows us to explore climatic differences and climate-related drivers. Additionally, it offers a chance to research how climatic variability affects the ecosystem and human health. By merging climate

and vegetation variation utilizing NDVI and EVI data to understand the relationship between vegetation and climate change under tropical climate conditions, it closes research gaps in Ghana. This study explores historical and projected vegetation and climate data, by sector, impacts, key vulnerabilities and what adaptation measures can be taken. It also explores the overview for a general context of how climate change is affecting Ghana.

1.7 Limitation Of The Study

The goal of time series modeling is to employ the simplest model feasible to account for as much data as possible while still developing an explanatory model of the data that does not over-fit the issue set.
Remote sensing data has additional limits that make this more difficult when dividing time series data into component pieces. It is almost certain that data from distant sensing will not provide the same level of precision.
Additionally, atmospheric factors can distort the visual findings, causing the vegetation's color to shift dramatically from image to image as a result of atmospheric factors (fog, ground moisture, cloud cover)

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Introduction

According to studies, there is now significant change in vegetation on the earth than there was thirty years ago, and it is distributed differently.

More than half of the changes they found are attributed to the consequences of a warmer climate, with people only being responsible for about a third. Perhaps surprisingly, they are unable to definitively link approximately 10% of the changes to either the climate or us.(**alex2013**)

Several models and hypotheses have been established in the environmental literature to explain the relationship between human behaviour, and environmental (forest) deforestation or depletion. Recent environmental and energy economics literature focuses on the energy consumption choices made by businesses and people in developing countries Gertler etal. (**gertler2016**). Africa's energy supply is made mainly of fuel wood and charcoal to a degree of about 58%. Specht et al. (**specht2015**) . Before other demands for forest goods like furniture and paper, the need for fuel wood for cooking and heating is frequently identified as the main driver of deforestation.

The causes of tropical forest decline are unclear, according to DeFries et al. (**defries2010**) . However, scientists were able to pinpoint the two primary causes of deforestation in the 21st century using information from satellite-based estimations in 41 different countries. The authors found a favorable association between forest loss and increases in urban population as well as agricultural exports using two methods of regression analysis. The same proof, however, was not discovered in the case of the increase in rural population. This suggests that forest loss is unavoidable in regions with high levels of human activity. Sedjo and Sohngen (1998) assessed various causes of climate-related forest degradation. Change and its effects on society and the economy. They discovered a positive correlation between forests in general, climate change, and timber harvesting. This suggests that earlier Research results indicating serious

implications have inflated the risk. They also assert that Concern exists over how climate change will affect the ecological values of forests. particularly if climate change occurs relatively gradually and its response is improved. Using a single DGVM and meteorological data from 15 distinct climate models for low and high emissions, Salazar et al. (2007) investigated the impact of climate change on the Amazon forest. They discovered from the models that rising temperatures are sufficient to induce the loss of forest and the conversion of the Amazon forest to savannah, even with high rainfall. This is true despite the wide range of expected precipitation changes over the Amazon forest. However, given that several DGVMs produce varied results, there are issues with effectiveness, consistency, and dependability because just one DGVM was used in their study. According to a related study, Huntingford et al. (2008) predicted that the Amazonian rain forest would see some attrition in the 21st century as a result of climate change.They also noted that the expected variations in temperature and rainfall have an impact on how much attrition will occur. The scientists did stress that because of a lag in the reaction to climate change, the decrease in the forest will be worse than most forecasts.

Mean while DGVMs have been used in other empirical investigations to understand climate change and its effects on forests. The method replicates competition between various vegetation kinds and forecasts potential changes in wooded regions due to a warming climate. To understand the mechanisms underlying changes in vegetation types and cover, Reu et al. (2011) investigated the link between climatic change and plant physiological processes. They discovered that when the climate warms, the concentration of species declines in the tropics but grows in the mid-latitudes.

The effects of temperature and precipitation fluctuations on humidity in forests are significant. The efficiency with which plants use water can be impacted by increasing water losses via evaporation and evapotranspiration, according to Mortsch (2006). When a warm temperature persists for a prolonged amount of time—for example, over a drought develops, significant moisture stress occurs. Depending on the characteristics of the forest, such as the type of habitat for fauna and flora, soil depth, and soil type, this process results in a decline in the development and health of trees.

CHAPTER THREE

3.0 METHODOLOGY

3.1 Introduction

The techniques used to study our model are outlined in this chapter, together with a comprehensive explanation of the mathematical tools and constructs, theorems, lemmas, and their justifications. The planning and execution processes of our technique make use of the R programming language. During the planning process, we considered where to get our data from and the procedures needed to build a time series model from the satellite data. We categorize our research as taking a quantitative approach. This research is specifically a causal-comparative experimental study with the goal of identifying the variability climate is responsible for in vegetation loss in Ghana.

3.2 Study Area

The Republic of Ghana, a nation in West Africa, will serve as the location for the experimental plots for this study. It shares borders with the Ivory Coast in the west, Burkina Faso in the north, and Togo in the east. It borders the Gulf of Guinea and the Atlantic Ocean to the south. Ghana's total size is 238,535 km2 (92,099 sq mi), and it is made up of a variety of biomes, from tropical rainforests to coastal savannas. Ghana, which has a population of over 31 million, is the second-most populous nation in West Africa, behind Nigeria.Accra, the nation's capital and largest city, as well as Kumasi, Tamale, and Sekondi-Takoradi, are other important cities

3.3 DATA

Data gathering. The majority of the quantitative data came from the Ghana Meteorological Agency and the Health Service. Rainfall, the highest temperature, and relative humidity are some of the meteorological data that were measured from 2010 to 2015 on a mean monthly basis. Additionally, based on monthly

incidences for the Kumasi Metropolitan Area from 2000 to 2022, the data for EVI were obtained.

3.3.1 Data Description and Inspection

Data from a time series is a set of observations made in a particular order over a period of time. There is a chance for correlation between observations because time series data points are gathered at close intervals. To help machine learning classifiers work with time series data, we provide several new tools. We first contend that local features or patterns in time series can be found and combined to address challenges involving time-series categorization. Then, a method to discover patterns that are helpful for classification is suggested. We combine these patterns to create computable categorization rules. In order to mask low-quality pixels, we will first collect data from Google Earth Engine in order to choose EVI values and Climate Change data.

Instead of analyzing the imagery directly, we will summarize the mean Climate values. This will shorten the analysis time while still providing an attractive and useful map. We will apply a smoothing strategy using an ARIMA function to fix the situation where some cells may not have EVI for a particular month. Once NA values have been eliminated, the time series will be divided to eliminate seasonality before the normalized data is fitted using a linear model. We will go to classify our data on the map and map it after we have extracted the linear trend.

Here, we made sure that we checked our data set for missing values and discovered there were none. The dataset's number of columns is shown in the table below, which also details the counts and datatype of various features in our dataset. No values are missing when there is a count of 133990. Anything below that denotes the existence of unreported values.

3.3.2 Time Series Forecasting Using Stochastic Models

The selection of a proper model is extremely important as it reflects the underlying structure of the series and this fitted model in turn is used for future forecasting. A time series model is said to be linear or non-linear depending on whether the current value of the series is a linear or non-linear function of past observations.

In general models for time series data can have many forms and represent different stochastic processes. There are two widely used linear time series models in literature.

Autoregressive (AR) and *Moving Average (MA)* models, combining these two, the Autoregressive Moving Average (ARMA) and *Autoregressive Integrated Moving Average (ARIMA)* models have been proposed in many literature. The *Autoregressive Fractionally Integrated Moving Average (ARFIMA)* model generalizes ARMA and ARIMA models. For seasonal time series forecasting, a variation of ARIMA. The *Seasonal Autoregressive Integrated Moving Average (SARIMA)* model is used.

ARIMA model and its different variations are based on the famous Box-Jenkins principle Hipel and McLeod (**hipel1994**) and so these are also broadly known as the Box-Jenkins models. **Vector Autoregression Model**

Model of vector autoregression. In econometrics, the Sims' vector autoregression (VAR) model has been widely employed to analyze multivariate time series. It provides better forecasting abilities than a univariate time series model and is a logical extension of the univariate autoregressive model to dynamic multivariate time series. Additionally, it identifies how each endogenous variable responds over time to a shock in both its own value and in every other variable. This method also enables the researcher to follow the data. The VAR model of order p Sims (**sims1980macroeconomics**) proposed has the following basic form:

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdot + A_p y_{t-p} + CD_t + \mu \tag{3.1}$$

Where $y_t = (y_{1t}, y_{2t}, ..., y_{kt})'$ is a vector of K observable endogenous variables. For the purposes of this study, yt = $(EVI_t, Temperature_t, Rain, Precipitation_t, Drought_t, Evaporation_t)'$ where EVI denotes the value of vegetation conditions each month, Temperature for both TMax and TMin, Rain denotes the amount of precipitation (mm), and Drought denotes the relative drought (%). All deterministic variables, including constants, linear trends, and seasonal dummy variables, are contained in D_t. An unobservable zero-mean white noise process in K dimensions, μ_t has a positive definite covariance matrix $E(\mu_t, \mu_t^{'}) = \sum_{\mu}$. We apply different limits on the parameter matrices A_i and C, which are of an appropriate dimension. Generalized least squares is used to estimate the model's parameters.

Augmented Dickey Fuller Test. For stationarity, the Augmented Dickey Fuller **(ADF)** test is a unit root test. The alternative hypothesis varies slightly depending on the equation used, whereas the null hypothesis for the ADF test is that there is a unit root. The time series is steady, which is the most straightforward

alternate theory. In time series analysis, unit roots can lead to unexpected outcomes.

$$\Delta y_t = \gamma y_{t-1} + \underbrace{\sum_{i=2}^{p} \beta_i \Delta y_{t-i+1}}_{\text{control for serial correlation}} + \epsilon_t \tag{3.2}$$

$$\rightarrow (\tau 1) \quad H_0 : \gamma = 0 \tag{3.3}$$

$$\tag{3.4}$$

$$\Delta y_t = \gamma y_{t-1} + \underbrace{a_0}_{\text{constant}} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \epsilon_t \tag{3.5}$$

$$\rightarrow (\phi 1) \quad H_0 : \gamma = 0 \quad \& \quad a_0 = 0 \tag{3.6}$$

$$\rightarrow (\tau 2) \quad H_0 : \gamma = 0 \tag{3.7}$$

$$\Delta y_t = \gamma y_{t-1} + a_0 + \underbrace{a_2 t}_{trend} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \epsilon_t \tag{3.8}$$

$$\rightarrow (\phi 2) \quad H_0 : \gamma = 0 \quad \& \quad a_0 = 0 \quad \& \quad a_2 = 0 \tag{3.9}$$

$$\rightarrow (\phi 3) \quad H_0 : \gamma = 0 \quad \& \quad a_0 = 0 \tag{3.10}$$

$$\rightarrow (\tau 3) \quad H_0 : \gamma = 0 \tag{3.11}$$

Optimal Lag Length Selection Criteria. OLS is used to estimate each of the system's individual equations. One of the following Information Criteria is minimized to determine the best lag order that is choosing optimal lag to reduce residual correlation

Criteria	Formular
AIC (n)	$= \log\det\left(\sum_{\mu}(n)\right) + \frac{2}{T} nK^2$
HQ(n)	$= \log\det\left(\sum_{\mu}(n)\right) + \frac{2logT}{T} nK^2$
SC(n)	$= \log\det\left(\sum_{\mu}(n)\right) + \frac{logT}{T} nK^2$
FPE(n)	$= \left(\frac{T+n^*}{T-n^*}\right)\det\left(\sum_{\mu}(n)\right)$

where $\sum_{\mu}(n)$ is the estimated by $T^{-1}\sum_{t=1}^{T}$

3.4 Structural Analysis

There are frequently many coefficients to comprehend, despite the fact that VAR coefficients describe the projected impact of a variable. Examining the model's residuals, which represent unforeseen contemporaneous events, is typically more prevalent. The next subsections explain some of the typical methods used for structural analysis of VAR models in a manner that is comparatively nontechnical.

3.4.1 Causality Analysis

Both the Granger-causality and instantaneous causality were investigated. For both tests, the vector of endogenous variables is divided into two subvectors,y_{1t} and y_{2t} with dimensions K_1 and K_2,ctively, so that $K = K_1 + K_2$ The subvector y_{1t} said to be Granger-causal for y_{2t} if the past of y_{1t} significantly helps predicting the future of y_{2t} a the past of y_{1t} one (**granger1969investigating**). For testing this property, a model of the form

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix} \alpha_{11,i} & \alpha_{12,i} \\ \alpha_{21,i} & \alpha_{22,i} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \end{bmatrix} + CD_t + \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} \tag{3.12}$$

Where, y_{1t} is not considered as Granger-causal for y_{2t} if and only if $\alpha_{21,i} = 0$, i = 1,2,...p. Therefore, this null hypothesis is tested that at least one of the $\alpha_{21,i}$, has a positive value. The constraints are tested using an F-test statistic with a distribution of $F(pK_1K_2, KT - n^*)$. Here, n^* is the entire number of system parameters, including those for the deterministic term (**lutkepohl2005new**). To test Granger-cause from y_{2t} to y_{1t}, the roles of y_{1t} and y_{2t} can be switched.In other words, Granger causality responds to the question of whether previous values of the variable x may help anticipate future values of the variable y.
Instantaneous causality is defined as having a nonzero correlation between μ_{1t} and μ_{2t}. Consequently, the null hypothesis is tested against the alternative

$$H_0 : E\left(\mu_{1t}, \mu_{2t}^{'}\right) \tag{3.13}$$

is checked for instantaneous causation versus the alternative of nonzero covariance between the two error vectors. This hypothesis is investigated using a Wald test statistic. **Analysis of impulse responses**: Exogenous and deterministic variables are viewed as fixed in impulse response analysis and can thus be removed from the system. Now, y_t stands for the adjusted endogenous variables. The process y_t has a Wold

moving average (MA) representation if it is stationary **(I(0))**, which is represented by

$$y_t = \Phi_0 \mu_t + \Phi_1 \mu_{t-1} + \Phi_2 \mu_{t-2} + \dots \tag{3.14}$$

Where $\Phi_0 = I_K$ and the Φ_s can be computed recursively as

$$\Phi_s = \sum_{j=i}^{s} \Phi_{s-j} A_j, \quad s = 1,2,\dots, \tag{3.15}$$

With $\Phi = I_K$ and $A_j = 0$ for $j > p$. The coefficients of this representation may be interpreted as reflecting the responses to impulses hitting the system. The *(i,j)th* element of the matrices Φ_s, regarded as a function of s, trace out the expected response of $y_{i,t+s}$ to a unit change in y_{it} holding constant all past values of y_t

3.5 Forecast Error Variance Decomposition

Denoting the $(i,j)th$ element of the orthogonalized impulse response coefcient matrix θ_n by $\theta_{ij,n}$, the variance of the forecast error $\left(y_{k,T+h} - y_{k,T+h|T}\right)$ is

$$\begin{aligned} \sigma_k^2(h) &= \sum_{n=0}^{h-1} \left(\theta_{k1,n}^2 + \cdot + \theta_{kK,n}^2\right) \\ &= \sum_{j=1}^{K} \left(\theta_{kj,0}^2 + \cdot + \theta_{kj,h=1}^2\right) \end{aligned} \tag{3.16}$$

The term $\left(\theta_{kj,0}^2 + \cdot + \theta_{kj,h=1}^2\right)$ is interpreted as the contribution of variable j to the h-step forecast error variance of variable k. Dividing the above terms by $\sigma_k^2(h)$ gives the percentage j to the h-step forecast error variance of variable k,

$$\omega_{kj}(h) = \frac{\left(\theta_{kj,0}^2 + \cdot + \theta_{kj,h=1}^2\right)}{\sigma_k^2(h)} \tag{3.17}$$

3.6 Forecast Performance Measures

While applying a particular model to some real or simulated time series, first the raw data is divided into two parts (**Training Set and Test Set**). The observations in the training set are used for constructing the desired model. Often a small sub-part of the training set is kept for validation purpose and is known as the **Validation Set**. Sometimes a preprocessing is done by normalizing the data or taking logarithmic or other

transforms. One such famous technique is the Box-Cox Transformation [23]. Once a model is constructed, it is used for generating forecasts. The test set observations are kept for verifying how accurate the fitted model performed in forecasting these values. If necessary, an inverse transformation is applied on the forecast values to convert them in original scale. In order to judge the forecasting accuracy of a particular model or for evaluating and comparing different models, their relative performance on the test dataset is considered.

Due to the fundamental importance of time series forecasting in many practical situations, proper care should be taken while selecting a particular model. For this reason, various performance measures are proposed in literature to estimate forecast accuracy and to compare different models. These are also known as performance metrics [24]. Each of these measures is a function of the actual and forecast values of the time series.

MAE	RMSE	MSE	MAPE
$\frac{1}{n}\sum_{t=1}^{n}\lvert e_t\rvert$	$\sqrt{MSE}=\sqrt{\frac{1}{n}\sum_{t=1}^{n}e_t^2}$	$\frac{1}{n}\sum_{t=1}^{n}e_t^2$	$\frac{1}{n}\sum\left(\frac{e_t}{y_t}\right)$

3.6.1 Description of Various Forecast Performance Measures

MAPE

This measure represents the percentage of average absolute error occurred. It is independent of the scale of measurement, but affected by data transformation.It does not show the direction of error. MAPE does not penalize extreme deviations. In this measure, opposite signed errors do not offset each other.

MSE

It is a measure of average squared deviation of forecast values. As here the opposite signed errors do not offset one another, MSE gives an overall idea of the error occurred during forecasting. It penalizes extreme errors occurred while forecasting. MSE emphasizes the fact that the total forecast error is in fact much affected by large individual errors, i.e. large errors are much expensive than small errors. MSE does not provide any idea about the direction of overall error. MSE is sensitive to the change of scale and data transformations. Although MSE is a good measure of overall forecast error, but it is not as intuitive and easily interpretable as the other measures discussed before.

RMSE

RMSE is nothing but the square root of calculated MSE. All the properties of MSE hold for RMSE as well.

MAD

measures the average absolute deviation of forecast values from original ones.It shows the magnitude of overall error, occurred due to forecasting.In MAE, the effects of positive and negative errors do not cancel out.Unlike MFE, MAE does not provide any idea about the direction of errors. For a good forecast, the obtained MAE should be as small as possible. Like MFE, MAE also depends on the scale of measurement and data transformations.Extreme forecast errors are not panelized by MAE.

Where: In each of the forthcoming definitions, y_t is the actual value,f_t is the forecast value, $e_t = y_t - f_t$ is the forecast error and n is the size of the test set. Also, $\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t$ is the test mean and

$\sigma^2 = \frac{1}{n-1}\sum_{t=1}^{n}(y_t - \bar{y})^2$is the test variance.

CHAPTER FOUR

4.0 DATA ANALYSIS AND RESULTS

4.1 Introduction

In this chapter, we present the results of the data analysis. We also interpret and discuss the results.

4.1.1 Preliminary Results

The study's data set includes time series of monthly EVI cases for the study area as well as various climatic factors like precipitation,evapiration, the temperature (Tmin and Tmax), drought and relative humidity . After extracting all the data from Google Earth Engine, there was a chance for correlation between observations because time series data points are gathered at close intervals. Instead of analyzing the imagery directly, we summarize the data through descriptive statistics for the five variables under consideration are shown in Table 1. The numbers for EVI are 36047, 87765, and 65318, respectively, which are the minimum, maximum, and average levels. In terms of climatic factors, the minimum, maximum, and average values of rainfall are 0.00mm, 379.80mm, and 115.26mm, respectively. In contrast, the minimum, maximum, and average values of maximum temperature are 27.40C, 36C, and 31.62C, respectively. The original data's time series plots are shown in Figure 2, which clearly shows that both EVI and the climatic variables under consideration are seasonal.

Figure 3's plot of the first difference in the data provides evidence of stationarity in both the climate Change and EVI variables.. This will shorten the analysis time while still providing an attractive and useful map. We will apply a smoothing strategy using an ARIMA function to fix the situation where some cells may not have EVI for a particular month. Once NA values have been eliminated, the time series will be divided to eliminate seasonality before the normalized data is fitted using a linear model. We will go to classify our data on the map and map it after we have extracted the linear trend. The To help machine learning classifiers work with time series data, we provide several new tools. We first contend that local features or

Table 4.1: Data Collected from 2000-2022 on Google Earth Engine

	Vegetation Indices	Climate Change				
date	EVI	Precipitation	Evapiration	TempMin	TempMax	Drought
2000-01-01	-0.38280560	1.0344619	2.936857e-05	217.1814	315.8379	-86.12496
2000-02-01	-0.08346374	0.9917288	2.634354e-05	210.7751	325.4743	-168.35120
2000-03-01	0.46029681	3.2132809	2.927312e-05	232.6620	332.0389	-220.69199
2000-04-01	0.68380747	5.1810527	4.515878e-05	224.0424	321.6810	-190.38286
2000-05-01	0.25644143	6.7150908	4.758059e-05	228.8765	318.3983	-210.67141
2000-06-01	-0.28410451	8.2797964	4.928518e-05	222.4945	295.6937	-209.06835

Table 4.2: Summary statistics for Climate Date and Vegetation Loss In Ghana

	Min.	1st Qu.	median	Mean	3rd Qu	Max.
EVI	0.3945	0.6423	0.7242	0.7168	0.8168	0.8672
Precipitation	0.1203	2.0311	3.5497	3.9427	5.4892	12.3576
Evapiration	1.682e-05	3.595e-05	4.215e-05	4.033e-05	4.560e-05	5.432e-05
TempMin	194.0	220.6	225.6	225.1	230.5	242.7
TempMax	274.4	294.1	313.8	309.0	322.2	348.2
Drought	-1354.42	-339.45	-228.47	-257.12	-89.24	369.64

patterns in time series can be found and combined to address challenges involving time-series categorization. Then, a method to discover patterns that are helpful for classification is suggested. We combine these patterns to create computable categorization rules. In order to mask low-quality pixels, we will first collect data from Google Earth Engine in order to choose EVI values and Climate Change data.

4.1.2 Selection of Variables

The purpose of this part is to examine how we got our answer and variables. First, we turned all categorical features into numerical attributes. Fortunaltely, some categroical features were already converted. Now, when we divided our variables into response and explanatory variables, some variables like dates and the unique policy number—were automatically discarded. Dates, policy number and other unrelevant variables which do not contribute to the project were removed, leaving us with these variables, PN, G, YOD, AOD, EL, MY, DUC, F, CAP, and NC

Again, we checked for the Pearson's correlation between our target variable's and the covariates' which are presented in the below

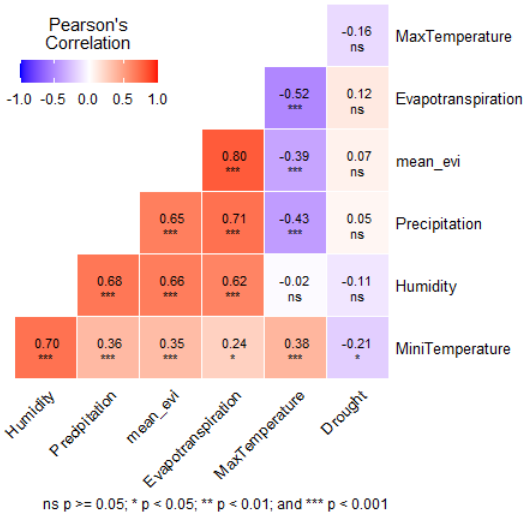

Figure 4.1:

4.1.3 Multicollinearity

Although we have known about our study's target variable, EVI, from the outset, picking the covariates was a challenge. The following variables, with the exception of EVI, were intended to be used as covariates until it was discovered that multicollinearity among the covariates increases the sum of squared error..We employed the Variance Inflation Factor (VIF) to test for multi-collinearity, and the findings is shown below,

Parameter	Values
Predictors	5
VIF	2.373
Condition number	7.939
Determinant	0.2260766568
Selected	Drought, TempMin, Precipitation, TempMax, Evapiration
Removed	Humidity

In order for our VIF score to fall between the ranges of 1 and 5, a piece of code was executed that automatically deleted variables with the highest VIF scores. The results of the chosen covariates are, This subsection statistically concluded that the variables Drought, TempMin, Precipitation, TempMax, and Evapiration will explain our target variable.

4.1.4 Stationarity and Differencing

Before building our models we first checked if the series is stationary.We used the R program to do the ADF test, and we looked at many ADF test statistics to determine whether a unit root existed. This is essential to

Table 4.3: Augmented Dickey Fuller (ADF) unit root test

	EVI			Precipitation		
	tau3	phi2	phi3	tau3	phi2	phi3
Test-Statistics	-12.23373	49.89733	74.84565	-12.90546	55.54404	83.31586
1pct	-3.98	6.15	8.34	-3.98	6.15	8.34
5pct	-3.42	4.71	6.30	-3.42	4.71	6.30
10pct	-3.13	4.05	5.36	-3.13	4.05	5.36
	TempMin			TempMax		
	tau3	phi2	phi3	tau3	phi2	phi3
Test-Statistics	-9.292711	28.810535	43.191412	-9.219737	28.363876	42.545576
1pct	-3.98	6.15	8.34	-3.98	6.15	8.34
5pct	-3.42	4.71	6.30	-3.42	4.71	6.30
10pct	-3.13	4.05	5.36	-3.13	4.05	5.36
	Evapiration			Drought		
	tau3	phi2	phi3	tau3	phi2	phi3
Test-Statistics	-11.37318	43.16267	64.74007	-3.421750	3.915871	5.862217
1pct	-3.98	6.15	8.34	-3.98	6.15	8.34
5pct	-3.42	4.71	6.30	-3.42	4.71	6.30
10pct	-3.13	4.05	5.36	-3.13	4.05	5.36

the VECM model and the VAR model in particular. That is, we needed to be determined that the time series is constant in mean and variance are constant and not dependent on time.Here, we look at a couple of methods for checking stationarity. If the time series is provided with seasonarity, a trend, or a change point in the mean or variance, then the influences need to be removed or accounted for. We use the Augmented Dickey–Fuller (ADF) t-statistic test to find if the series has a unit root (a series with a trend line will have a unit root and result in a large p-value).(**dickey1979distribution**)The test, shown in Tables 2(a) and 2(b), disproved the unit root hypothesis for all time series taken into account, suggesting that the relationships between the numerous variables investigated below are not fictitious.

4.2 VAR Estimation

With 12 lags for each variable in each equation, a VAR model of the monthly EVI and Climate Change variables was estimated. There are 4x12 unconstrained coefficients in each equation, plus one for a constant

Table 4.4: Optimal lag length selection

Lag	AIC(n)	HQ(n)	SC(n)	FPE(n)
1	-8.242332	-8.006335	-7.655762	0.0002633
2	-9.304816	-8.866535	-8.215470	0.0000910
3	-9.743066	-9.102502	-8.150946	0.0000588
4	-10.142686	-9.299839	-8.047791	0.0000395
5	-10.286777	-9.241647	-7.689107	0.0000343
6	-10.338471	-9.091058	-7.238027	0.0000328
7	-10.345081	-8.895385	-6.741861	0.0000328
8	-10.344697	-8.692718	-6.238702	0.0000331
9	-10.310857	-8.456595	-5.702089	0.0000347
10	-10.251454	-8.194908	-5.139910	0.0000374
11	-10.250272	-7.991444	-4.635954	0.0000382
12	-10.295256	-7.834144	-4.178163	0.0000374

Table 4.5: Granger causality tests.

Cause variable	Null hypothesis	F-value	p-value	Decision
Precipitation	does not Granger-cause	2.1563	0.01464	
Evapiration	does not Granger-cause	1.5398	0.1112	
TempMin	does not Granger-cause	3.0049	0.0006276	
TempMax	does not Granger-cause	2.7462	0.001685	
Drought	does not Granger-cause	0.9235	0.5241	

and one for a trend. Four tests were examine, the Final Prediction Error (FPE) test, the Hannan Quinne (HQ) test, the Information Criteria proposed by Akaike (AIC), and the Schwarz (SC) test were used to determine the appropriate number of lags. The results of all four tests suggested that up to 12 lagged monthly values would be sufficient. All the dynamics in the data are captured and employed in this analysis thanks to a lag duration of 12. (Table 3).

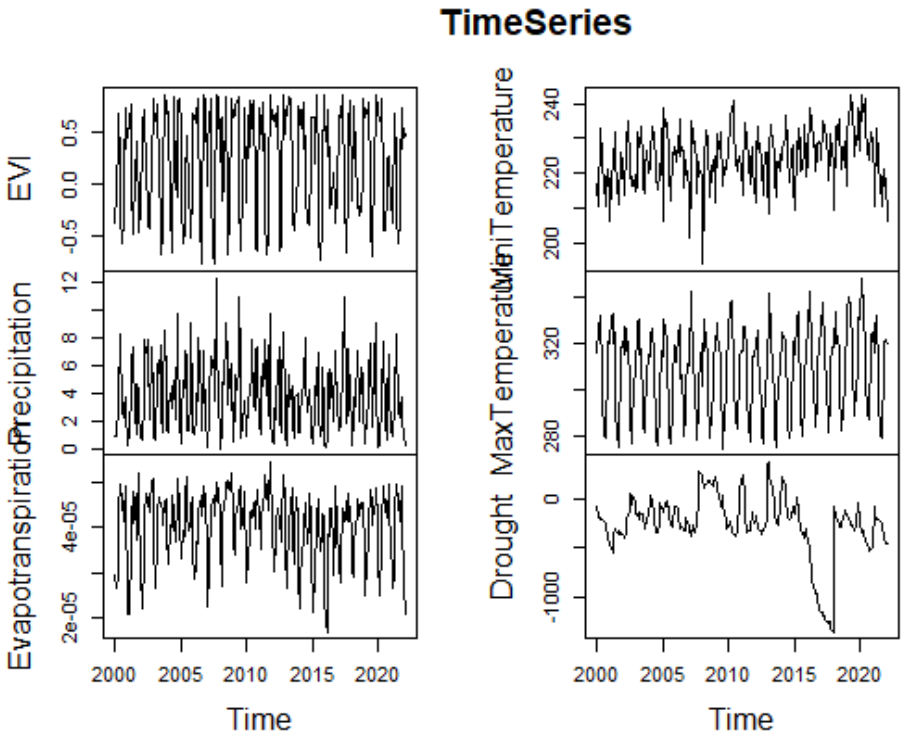

Figure 4.2:

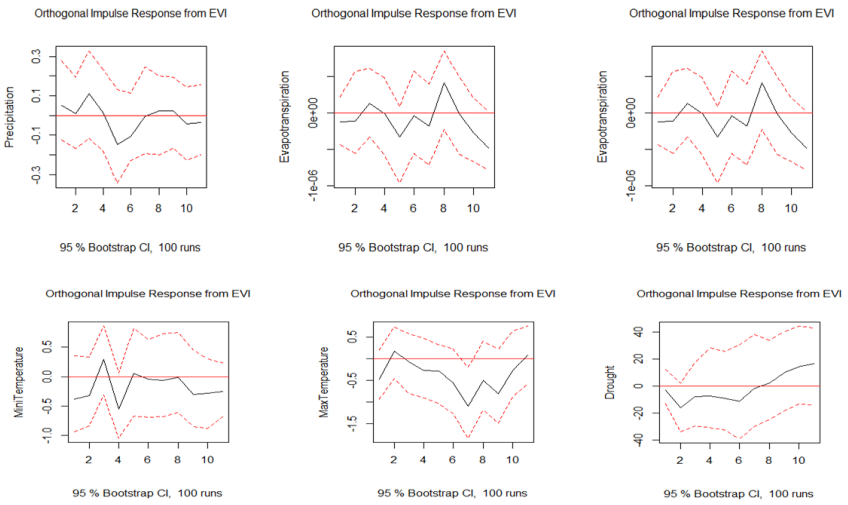

Figure 4.3:

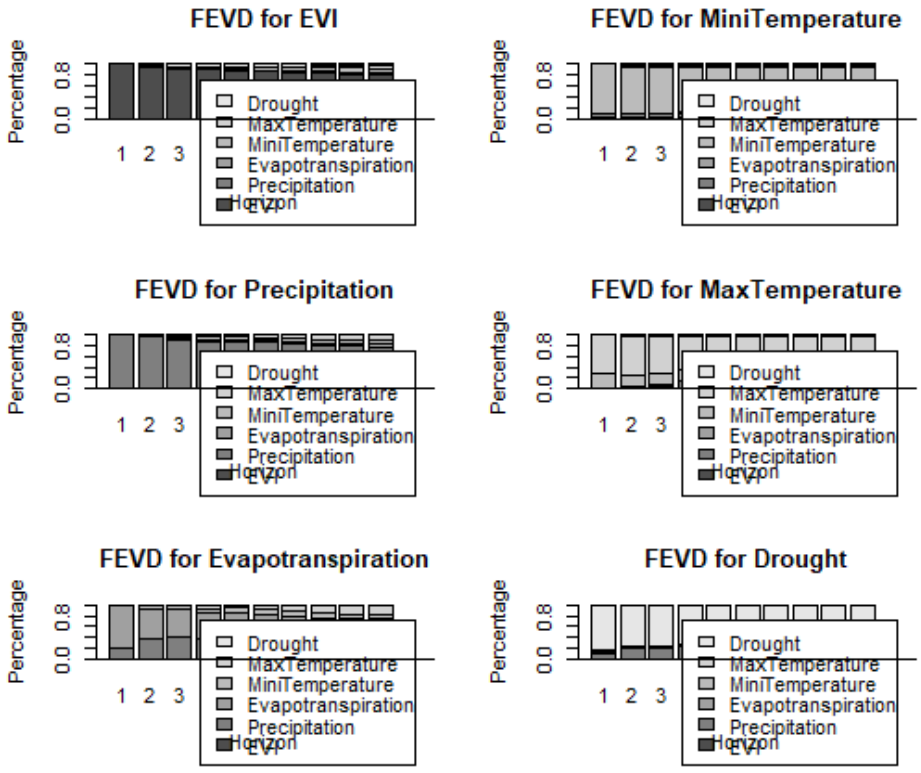

Figure 4.4:

4.3 Forecast for EVI

Future EVI cases can be predicted using the created VAR (12) model as a predictive model. Figure and Table both revert the forecasts for the number of differed EVI during the first half of 2022 to the original level. These predictions

4.4 Prediction accuracy

Regardless of whether the forecasting mistake is positive or negative, the mean absolute percentage error (MAPE) gives a general idea of the average size of forecasting error represented as a % of the actual observed number . The fitted model's MAPE is calculated to be 12.95% using (14) which suggests that its projections may be quite accurate .

CHAPTER FIVE

5.0 CONCLUSION & RECOMMENDATIONS

5.1 Introduction

This chapter contains the summary of our findings and the recommendations from our findings. These recommendations are necessary information for the Health Directorate in Sunyani and also for Mathematicians in the study of nonlinear systems.

5.2 Conclusion

Appendix Chapter

Appendix B Chapter