UNIVERSITY OF ENERGY AND NATURAL RESOURCES, SUNYANI

THE VARIABILITY CLIMATE CHANGE IS RESPONSIBLE FOR IN GHANA

KALONG BONIFACE

FUGAH SELETEY MITCHELL

DEPARTMENT OF MATHEMATICS AND STATISTIC

SCHOOL OF SCIENCE

THE VARIABILITY CLIMATE CHANGE IS RESPONSIBLE FOR IN GHANA

by

KALONG BONIFACE UEB3603118

FUGAH SELETEY MITCHELL UEB3602818

B.Sc. Actuarial Science

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 Actuarial Science

DECLARATION AND CERTIFICATION

We hereby Certified that the thesis entitled "**THE VARIABILITY CLIMATE CHANGE IS RESPONSIBLE FOR IN GHANA**", submitted by **Kalong Boniface** and **Fugah Seletey Mitchell** to the **DEPARTMENT** , for the award of Bachelor's degree has been accepted by the external examiners and that we have successfully defended the thesis held today.This dissertation is the result of our own independent investigation and research, we now announce. We are solely liable for any errors in this work even if it has not been submitted anyplace as a long essay or thesis with the intent of awarding a degree.

Supervisor's Certification

This study was carried out under the supervisory committee in accordance with the guidelines on supervisions of graduate studies.

KALONG BONIFACE Student	.... Signature	.... Date
FUGAH SELETEY MITCHELL Student	.... Signature	.... Date
Certified by: MR. JUSTICE AMENYOR KEESIE Supervisor	.... Signature	.... Date
Certified by: DR. ELVIS KWABENA DONKOH Head of Department	.... Signature	.... Date

ABSTRACT

In addition to the environmental damage that illicit mining causes, hundreds of young men have died after becoming trapped in unlawful mining holes while looking for minerals.

This study aimed to quantify, map, analyse vegetation cover distributions and changes across southern part of Ghana,and to model the effect of climatic variability on vegetation and using vector autoregression (VAR) models and if possible assign the residual errors to galamsey. Monthly weather information for 2000 (precipitation, relative drought,evaporation, both maximum and minimum temperature). Data on climatic variables and information on EVI were obtain from the Google Earth Engine from 2000 to 2021 time period. The Granger and immediate causality tests' findings indicated that all factors influence the spread of EVI. The results of the impulse response analyses showed that the months of September, March, and October, respectively, saw the strongest favorable effects of maximum temperature, drought, and precipitation on EVI. As much as 12.65% of the predicted variance shows a different degree of EVI dependence on climatic variables. The decomposition of forecast variance indicates a small varying degree of EVI dependence on the climatic variables, with as high as 12.65% of the variability in the trend of EVI which has been explained by past innovations in maximum temperature alone. This is quite significant and therefore, policy-makers should not ignore temperature when formulating policies to address vegetation loss

DEDICATION

This study is devoted wholeheartedly to our dear parents, who have served as our source of motivation and strength when we had considered giving up. They also continuously support us financially, emotionally, spiritually, morally, and emotionally. To all of our brothers, sisters, family members, mentors, friends, and classmates who offered words of wisdom and motivation to complete our study. We also dedicate this work to the All-Powerful God. We are grateful for your direction, fortitude, mental power, protection, abilities, and for providing us with a long and healthy life. We present all of these to you.

ACKNOWLEDGMENTS

First and foremost, we give praise to the Lord, the Almighty, for his direction, power, and wisdom. Next, we sincerely appreciate our supervisor, Mr. Justice Amenyo Kessie, for his constant support, encouragement, and the time he spent reading this research paper, critiquing it as necessary, explaining the criticism for our understanding, and offering immensely helpful suggestions and recommendations on how to structure this work.

Contents

List of Tables

List of Figures

List of Abbreviation

AR Autoregression

MA Moving Average

ARMA Autoregresive Moving Average

ARIMA Autoregressive Integrated Moving Average

AFIMA Autoregressive Fractionally Integrated Moving Average

VAR Vector Autoregression Model

EVI Enhance Vegetation Index

TMin and TMax Minimum and Maximum Temperature

ADF Augmented Dickey Fuller

OLS Optimal Lag Length Selection Criteria

AIC Akaike Information Criterion

HQ Hannan-Quinn criterion

SC Schwarz Criterion

FPE Final Prediction Error criterion

FEVD Forecast Error Variance Decomposition

CHAPTER ONE

1.0 INTRODUCTION

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",(Mantey et al., 2017) 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 (Gracia, 2018). 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 (News, n.d.).

Illegal mining causes damage to the land and water supply (Ansah, 2017) . 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 (Allotey, 2017) . 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, n.d.).

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 (Barenblitt et al., 2021) . 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.

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.

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.

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 loss and Climate change in Ghana?

Research Objectives

The purpose is to establish an understanding of 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.Unfortunately, the causes of deforestation and forest degradation have not been adequately defined in the environmental literature. According to

hosonuma2012assessment , there are four causes of forest degradation: logging for wood, uncontrolled fires, livestock grazing in forests, and five deforestation drivers, and fuel (wood/charcoal) (commercial agriculture, subsistence agriculture, mining, infrastructure and urban expansion). According to these drivers, this study reclassifies the drivers for an empirical examination into human conduct or activity and climatic change by;

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

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 climatic and vegetation (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.

Limitation 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 climatic and vegetation data to understand the relationship between vegetation and climate change closes research gaps in Ghana. This study explores historical and projected vegetation and climatic 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's Vegetation .

CHAPTER TWO

2.0 LITERATURE REVIEW

Introduction

According to studies, there is now a 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 behavior, 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 (**gertler2016**). Africa's energy supply is made mainly of fuel wood and charcoal to a degree of about 58%.(**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 the rural population. This suggests that forest loss is unavoidable in regions with high levels of human activity. (Sohngen & Sedjo, 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, (Malhi et al., 2008) 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, (Sitch et al., 2008) predicted that the Amazonian rainforest 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.

Meanwhile, 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, (Kattge 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 the humidity in forests are significant. The efficiency with which plants use water can be impacted by increasing water losses via evaporation, 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. The forest transition theory outlines the process through which industrialization and urbanization modify forest cover (Rudel et al., 2010). Industrialization and urbanization are the results of economic progress, which drives the active people from rural to urban areas. Early economic growth is characterized by a large percentage of forested land and a low pace of deforestation. The amount of forest cover is decreasing due to increased deforestation at the middle stage. A more advanced level of development causes the rate of deforestation to slow down, eventually stabilizing and restoring the forest cover. Human population density, level of development, economic structure, pressures of the global economy, and governmental policies all have an impact on this pattern.

CHAPTER THREE

3.0 METHODOLOGY

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.

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 biomass, from tropical rain forests 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

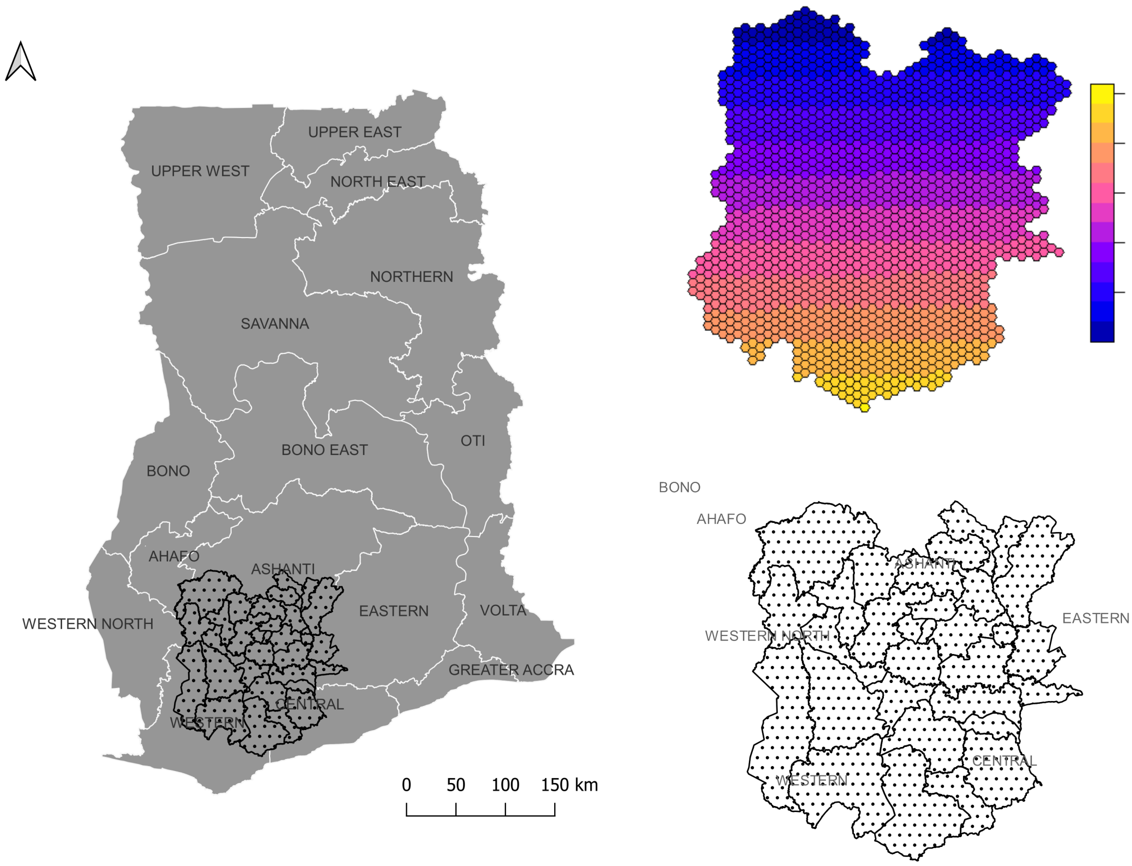

Figure 1: Study Area

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.

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.

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 (Sims, 1980) proposed has the following basic form:

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

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 CD_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, detecting the presence of a unit root in quite general time series models. Our approach is nonparametric with respect to nuisance parameters and thereby allows for a very wide class of weakly dependent and possibly heterogeneously distributed data. The tests accommodate models with a fitted drift and a time trend so that they may be used to discriminate between unit root nonstationarity and stationarity about a deterministic trend. The limiting distributions of the statistics are obtained under both the unit root null and a sequence of local alternatives. The latter noncentral distribution theory yields local asymptotic power functions for the tests and facilitates comparisons with alternative procedures due to Dickey & Fuller (Phillips & Perron, 1988).

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

$$\rightarrow (\tau 1) \quad H_0 : \gamma = 0 \tag{3}$$

$$\tag{4}$$

$$\Delta y_t = \gamma y_{t-1} + \underbrace{a_0}_{\text{constant}} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \epsilon_t \tag{5}$$

$$\rightarrow (\phi 1) \quad H_0 : \gamma = 0 \quad \& \quad a_0 = 0 \tag{6}$$

$$\rightarrow (\tau 2) \quad H_0 : \gamma = 0 \tag{7}$$

$$\Delta y_t = \gamma y_{t-1} + a_0 + \underbrace{a_2 t}_{trend} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \epsilon_t \tag{8}$$

$$\rightarrow (\phi 2) \quad H_0 : \gamma = 0 \quad \& \quad a_0 = 0 \quad \& \quad a_2 = 0 \tag{9}$$

$$\rightarrow (\phi 3) \quad H_0 : \gamma = 0 \quad \& \quad a_0 = 0 \tag{10}$$

$$\rightarrow (\tau 3) \quad H_0 : \gamma = 0 \tag{11}$$

(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.

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
Akaike Information Criterion,AIC (n)	$= \log \det \left(\sum_{\mu}(n) \right) + \frac{2}{T} n K^2$
Hannan-Quinn criterion,HQ(n)	$= \log \det \left(\sum_{\mu}(n) \right) + \frac{2logT}{T} n K^2$
Schwarz Criterion,SC(n)	$= \log \det \left(\sum_{\mu}(n) \right) + \frac{logT}{T} n K^2$
Final Prediction Error criterion,FPE(n)	$= \left(\frac{T + n^*}{T - n^*} \right) \det \left(\sum_{\mu}(n) \right)$

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

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.

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 (Granger, 1969). For testing this property, a model of the form

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

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 (Lütkepohl, 2005). 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

$$H_0 : E\left(\mu_{1t}, \mu_{2t}^{'}\right) \tag{13}$$

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

$$y_t = \Phi_0 \mu_t + \Phi_1 \mu_{t-1} + \Phi_2 \mu_{t-2} + \dots \tag{14}$$

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

$$\Phi_s = \sum_{j=i}^{s} \Phi_{s-j} A_j, \quad s = 1, 2, \dots, \tag{15}$$

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

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

$$\begin{aligned} \sigma_k^2(h) &= \sum_{n=0}^{h-1} \left(\theta_{k1,n}^2 + \cdot + \theta_{kK,n}^2\right) \\ &= \sum_{j=1}^{K} \left(\theta_{kj,0}^2 + \cdot + \theta_{kj,h=1}^2\right) \end{aligned} \tag{16}$$

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,

$$\omega_{kj}(h) = \frac{\left(\theta_{kj,0}^2 + \cdot + \theta_{kj,h=1}^2\right)}{\sigma_k^2(h)} \tag{17}$$

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.

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

Table 1: Various Forecast Performance Measures

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

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

Introduction

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

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,evaporation, temperature (Tmin and Tmax), drought and drought . 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 six variables under consideration which are shown in Table . The numbers for EVI are 0.3945, 0.8672, and 0.7192, respectively, which are the minimum, maximum, and average levels. In terms of climatic factors, the minimum, maximum, and average values of Precipitation are 0.1203mm, 12.3576mm, and 3.9604mm, respectively. In contrast, the minimum, maximum, and average values of maximum temperature are 274.4C, 348.2C, and 309.3C, respectively. The original data's time series plots are shown in Figure 4, which clearly shows that both EVI and the climatic variables under consideration are seasonal.

Variable & Parameter Definition

Find in the tables the definition of variables and parameters respectively as used in this study.

Selection of Variables

The purpose of this part is to examine how we got our answer and variables. First, we extract all our climatic data using one of the coordinate in these cells . To shorten the analysis time while still

Table 2: Variable Definition.

Variable	Definition
EVI	Vegetation Index
Precipitation	Atmospheric water vapor
Evaporation	Actual evapotranspiration, derived using a one-dimensional soil water balance model
TempMin	Minimum Temperature
TempMax	Maximum Temperature
Drought	Palmer Drought Severity Index

Table 3: Data Collected from 2000-2022 on Google Earth Engine

	Vegetation Indices	Climate Change				
Date	EVI	Precipitation	Evaporation	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

providing an attractive and useful map. We apply a smoothing strategy using an ARIMA function to fix the situation where some cells do 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.

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.

Note that the Analysis was divided into two, the uni-variate and Multivariate Analysis, where we use the uni-variate analysis to show the foot print of vegetation loss and multivariate to analysis how climatic variable are also contributing to such events.In that case we mask low-quality pixels, by first collecting data from Google Earth Engine in order to choose EVI values and Climate Change data. The Plot 4 shows the first difference in the data provides evidence of stationarity in EVI variables.Which we used to create the map.

For mulitvariate analysis, we checked for the Pearson’s correlation between our target variable’s and

Table 4: 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
Evaporation	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

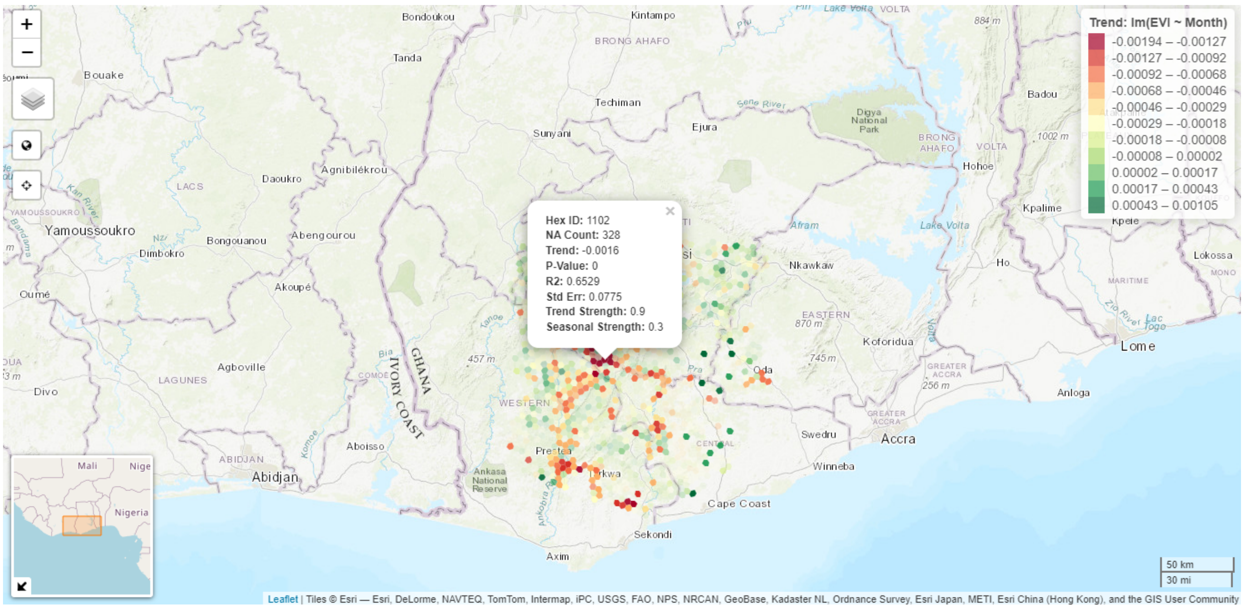

Figure 2: EVI Classification

the covariates' which are presented in the Figure 3 below

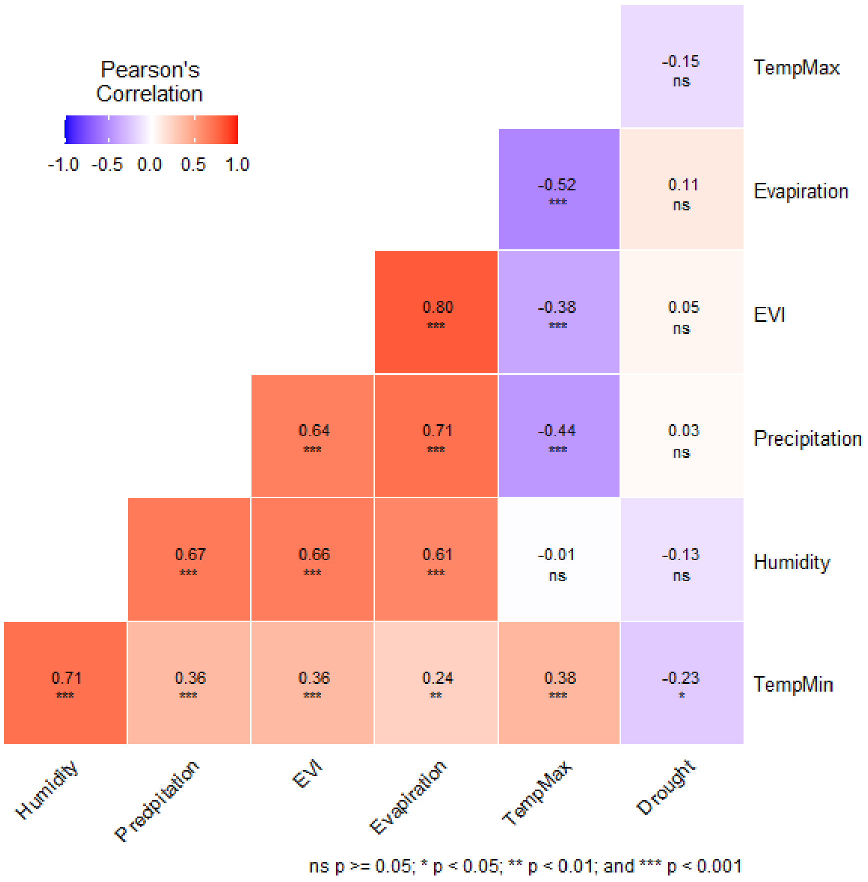

Figure 3: Pearson's Correlation Between Variables

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 in Table

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. We statistically concluded that the variables Drought, TempMin, Precipitation, TempMax, and Evaporation will explain our target variable.

Table 5: Best Variance Inflation Factor

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

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 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, 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)(Dickey & Fuller, 1979).The test, shown in Table , approved the unit root hypothesis for all time series taken into account, suggesting that the relationships between the numerous variables investigated below are not fictitious.

Table 6: 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
	Evaporation			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

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 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).

Table 7: 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

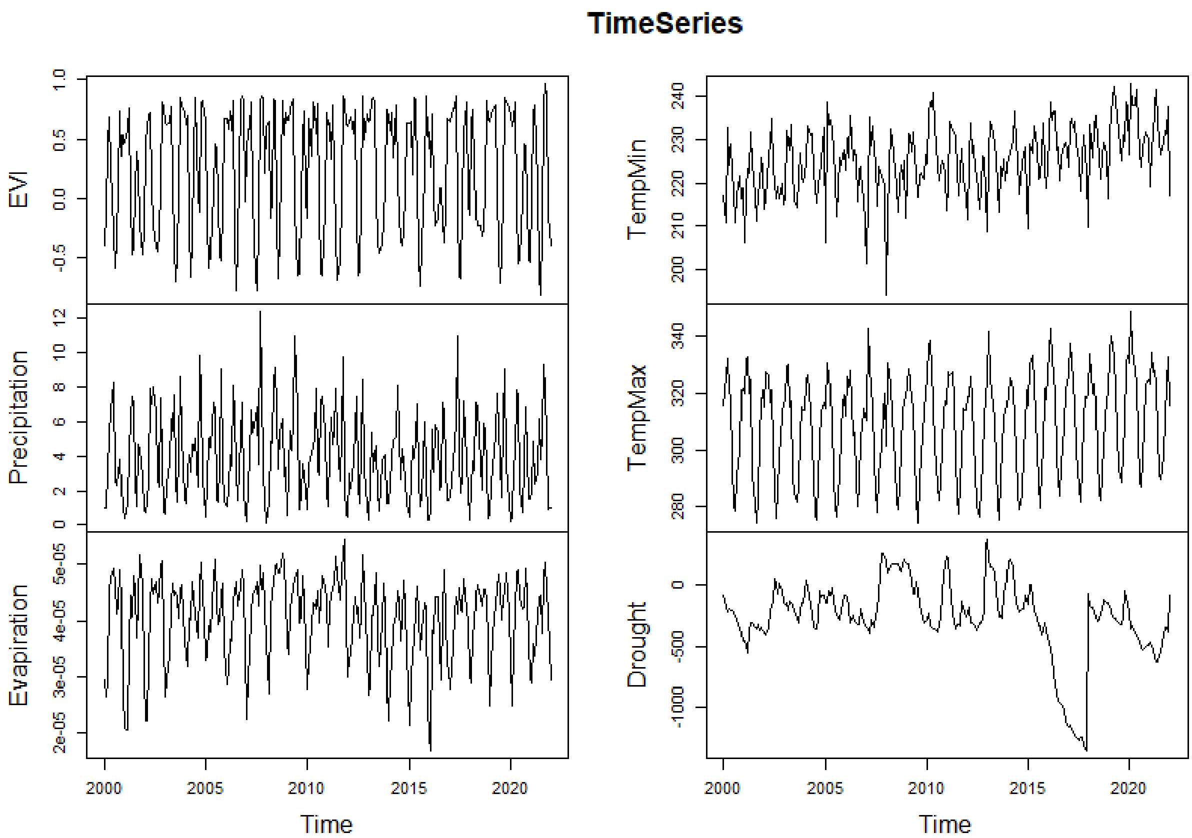

Figure 4: Time Series Plot of all Variables

Table 8: EVI forecast for the next 12 month.

Month	Forecast	Lower	Upper	CI
Feb	-0.07548865	-0.65614302	0.5051657	0.5806544
Mar	0.14803164	-0.56646127	0.8625245	0.7144929
Apr	0.69105704	-0.04392727	1.4260414	0.7349843
May	0.72499702	-0.05679272	1.5067868	0.7817897
Jun	0.44066359	-0.41564183	1.2969690	0.8563054
Jul	-0.11537243	-0.98453117	0.7537863	0.8691587
Aug	-0.37165805	-1.24946226	0.5061462	0.8778042
Sep	-0.16672258	-1.05855108	0.7251059	0.8918285
Oct	0.16015717	-0.73872579	1.0590401	0.8988830
Nov	0.37177202	-0.53806829	1.2816123	0.9098403

Table 9: Granger causality tests.

Cause variable	Null hypothesis	F-value	p-value	Decision
Precipitation	does not Granger-cause EVI	2.1563	0.01464	Reject the null hypothesis
Evaporation	does not Granger-cause EVI	1.5398	0.1112	Fail to Reject the null hypothesis
TempMin	does not Granger-cause EVI	3.0049	0.0006276	Reject the null hypothesis
TempMax	does not Granger-cause EVI	2.7462	0.001685	Reject the null hypothesis
Drought	does not Granger-cause EVI	0.9235	0.5241	Fail to Reject the null hypothesis

Impulse Response Analysis

The dynamic interactions between EVI and the Climatic variables of the VAR (12) process were examined using impulse response analysis. Figure 5 shows the orthogonal impulse response of EVI(Vegetation Condition) to the climate factors. The reaction of EVI exhibits a clear variation; the seventh month (July) has the largest negative effect of maximum temperature on EVI while we can see a rise in the Evaporation side and the third month has the lowest adverse effect (March). Additionally, the third month (March) is when drought has the most positive impact on EVI, whereas the tenth month (October) is when precipitation has the greatest good impact (October).

Decomposition of Variance

In order to evaluate VAR models, many people use forecast error variance decomposition (FEVD). Figure 6 displays the results for the FEVD for EVI and the climate factors. The findings show that, on average, past values EVI(Vegetation Condition) account for about 85.67% of the variability in the trend of EVI, while past innovations in maximum temperature account for a sizable percentage (around 5.65%) of the variability in the trend of EVI. Additionally, although earlier advances in Evaporation data have been able to account for around 3.10 percent of the variability in the trend of

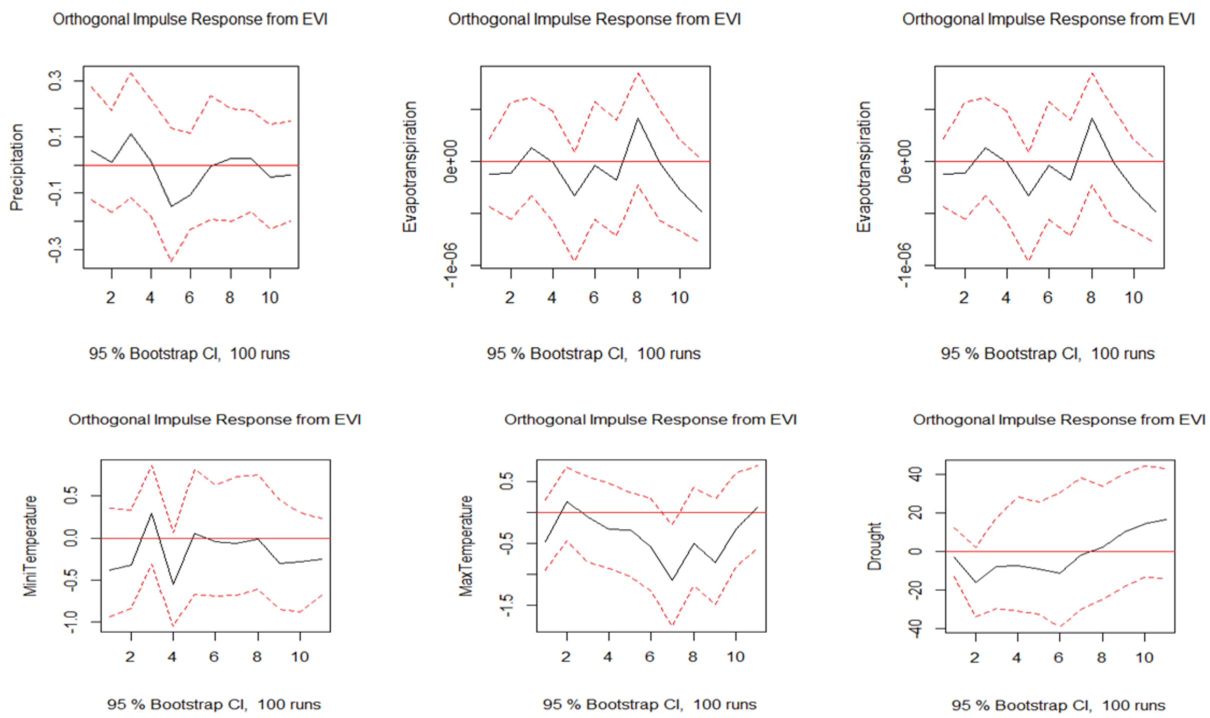

Figure 5: Impulse Response Analysis

Table 10: Forecast Accuracy on EVI Training set

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-1.384005e-17	2.665621e-01	2.065071e-01	-Inf	Inf	0.00131	0.0123

EVI, past innovations in relative Precipitation have only been able to account for 0.58 percent of that variability.

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

Prediction Accuracy

Regardless of whether the forecasting mistake is positive or negative, the mean absolute percentage error (MAE) gives a general idea of the average size of forecasting error represented as a % of the actual observed number .In this case the closer MAE is to zero (0) the more accurate the model is. The fitted model's MAE is calculated to be **2.065071e-01** using (10) which suggests that its projections may be quite accurate .

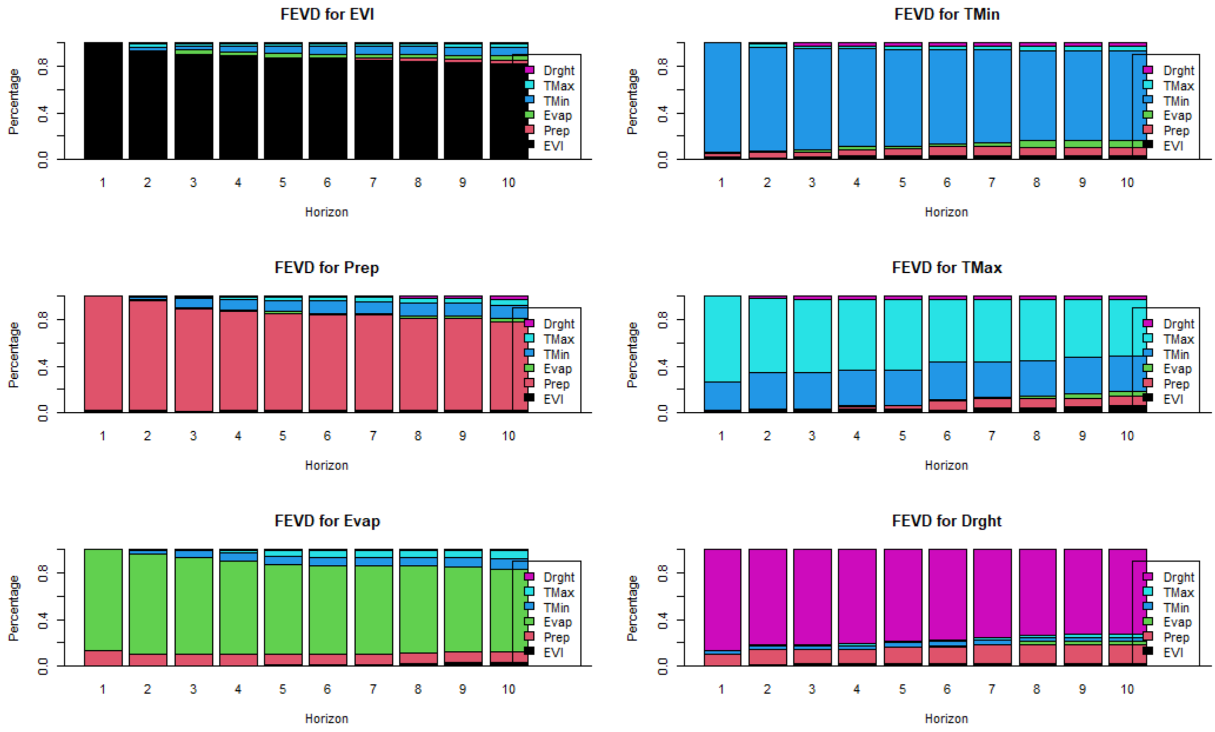

Figure 6: Forecast Error Variance Decomposition (FEVD)

The Granger and immediate causality tests reveal that only three climate factors have an impact on EVI. The impulse response analyses show that the ninth, third, and tenth months, respectively, had the strongest nagetive effects of maximum temperature, relative precipitation, and evaporation on EVI. With less than 20% of the variability in the trend of EVI being explained by historical innovations in Climate Change, the decomposition of predicted variance shows varied degrees of EVI dependence on climatic variables. Policymakers can utilize the study's findings to assist them develop policies by understanding how climatic variability affects the incidence of EVI in the Study Area.

Discussion

This study found that just three meteorological factors have an impact on EVI in the studied area,but it might deffer from other cells. Maximum temperature has the biggest beneficial impact on EVI in September and the lowest negative impact in March. Additionally, the biggest favorable effects of precipitation and dryness on EVI are recorded in March and October, respectively. Additional findings show that, while a larger amount of the variability in the trend of EVI may be attributed to past innovations in EVI instances, a sizable fraction (approximately 12.65%) of the trend of EVI variability can be attributed to past innovations in maximum temperature.Additionally, only 2.58 percent of the variability in the trend of EVI has been described by prior innovations in drought, compared to 11.10 percent explained by past innovations in precipitation data.

This further suggests that the three climatic factors have different effects on EVI, with maximum temperature having a bigger impact than precipitation, followed by drought, and so on. The forecast of EVI will therefore be improved by modeling EVI and the climatic variables jointly.

These results shown that, in contrast to the vast field cleared by either Galamseyers or farmers, relative drought, maximum temperature, and precipitation are farily responsible for determining the variability in terms of vegetation state. The differences in the results could have been caused by a variety of circumstances, including human action, the research site, and the length of the study (Galamsey). To predict future EVI instances, the established VAR (12) model can also be used as a predictive model. According to the calculated anticipated numbers, Galamseyers had a significant impact on the vegetation in this region of Ghana during the first half of 2016. One of the study's shortcomings is the absence of data over a longer time period.

CHAPTER FIVE
5.0 CONCLUSION & RECOMMENDATIONS

Introduction

This chapter contains the summary of our findings and the recommendations from our findings. These recommendations are necessary information for the Vegetation Changes in Ghana and also for Mathematicians in the study of Time Series systems.

Conclusion

This paper develops and estimates a Victor Autoregression (VAR) model of the monthly Vegetation condition and some important climatic variables including precipitation, maximum temperature, and relative drought in the southern part of Ghana. The model is used to investigate the dynamic link between vegetation and climatic variability. The model is also used to simulate the responses of EVI to innovations in climatic variability. Results of the Granger causality tests lead to a conclusion that EVI is influenced by only three climatic variables for that particular study area(Cell number 196). The impulse response analyses indicate that the highest positive effect of temperature, drought, and precipitation on EVI is observed in the ninth, third, and tenth months, respectively. The decomposition of forecast variance indicates varying degree of EVI dependence on the climatic variables, with as high as 12.65% of the variability in the trend of EVI being explained by past innovations in temperature alone. Results obtained from this study are useful for policy-makers as this will help come up with policies knowing the effects of climatic variability on EVI incidence in the Ghana.

RECOMMENDATIONS

Time series forecasting is a fast growing area of research and as such provides many scope for future works. One of them is the Combining Approach, i.e. to combine a number of different and dissimilar methods to improve forecast accuracy. A lot of works have been done towards this direction and various combining methods have been proposed in literature. Together with other analysis in time series forecasting, we have thought to find an efficient combining model, in future if possible. With the aim of further studies in time series modeling and forecasting.Therefore time series models, the rule of thumb is that one should have at least fifty (50) to sixty (60) data points but preferably more than hundred (100) observations (Box & Tiao, 1975). It is therefore suggested that future studies in this area of interest should consider more than hundred data points.. Therefore, it is recommended that future research in this area of interest take into account more than a hundred data points.Using satellite data to help inform reclamation projects. Knowing the location and extent of degraded forests can help land managers better project the labor and expense to reclaim an area (by planting tree seedlings or adding plants that could detoxify the area, for instance)

Bibliography

Allotey, G. A. (2017). Stop galamsey in 3 weeks or face the law - amewu. *Ghana News*. http://citifmonline.com/2017/03/29/stop-galamsey-in-3-weeks-or-face-the-law-amewu/

Ansah, M. E. (2017). Galamsey, pollution destroying water bodies in ghana - water company. *Ghana News*. http://www.leakxgh.com/2017/05/galamsey-pollution-destroying-water.html

Barenblitt, A., Payton, A., Lagomasino, D., Fatoyinbo, L., Asare, K., Aidoo, K., Pigott, H., Som, C. K., Smeets, L., Seidu, O., & Wood, D. (2021). The large footprint of small-scale artisanal gold mining in ghana. *Science of the Total Environment*, *781*. https://doi.org/10.1016/j.scitotenv.2021.146644

Box, G. E., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical association*, *70*(349), 70–79.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), 427–431.

Gracia, Z. (2018). Causes and effects of galamsey in ghana. *Yen.com.gh - Ghana news*. https://yen.com.gh/104844-causes-effects-galamsey-ghana.html

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424–438.

Gyekye, J. (n.d.). Md of ghana water company limited says fight against galamsey is being lost. *Ghana Broadcasting Corporation*. http://www.gbcghana.com/1.11923484

Kattge, J., Diaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P. B., Wright, I. J., et al. (2011). Try–a global database of plant traits. *Global change biology*, *17*(9), 2905–2935.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Malhi, Y., Roberts, J. T., Betts, R. A., Killeen, T. J., Li, W., & Nobre, C. A. (2008). Climate change, deforestation, and the fate of the amazon. *science*, *319*(5860), 169–172.

Mantey, J., Owusu-Nimo, F., Nyarko, K. B., & Aubynn, A. (2017). Operational dynamics of "galamsey" within eleven selected districts of western region of ghana. *Journal of Mining and Environment*, *8*, 11–34. https://doi.org/10.22044/jme.2016.627

Mortsch, L. (2006). Impact of climate change on agriculture, forestry and wetlands. *Climate change and managed ecosystems*, 45–67.

News, B. (n.d.). *Bbc news - women die in ghana mine collapse*. http://news.bbc.co.uk/2/hi/africa/8356343.stm

Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*(2), 335–346.

Rudel, T. K., Schneider, L., & Uriarte, M. (2010). Forest transitions: An introduction. *Land use policy*, *27*(2), 95–97.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1–48.

Sitch, S., Huntingford, C., Gedney, N., Levy, P., Lomas, M., Piao, S., Betts, R., Ciais, P., Cox, P., Friedlingstein, P., et al. (2008). Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five dynamic global vegetation models (dgvms). *Global change biology*, *14*(9), 2015–2039.

Sohngen, B., & Sedjo, R. (1998). A comparison of timber market models: Static simulation and optimal control approaches. *Forest Science*, *44*(1), 24–36.

Appendix Chapter 1

Data Extraction Using R code From Google Earth Engine

```
library(tidyverse)  # for data wrangling and visualization

library('sf')

library(tibble)

library(lubridate)

library(rgee)

library(reticulate)

ee_install()

ee_check()


ee_Initialize("kalong",drive = TRUE) # initialize GEE,

#this will have you log in to Google Drive
```

Load shape file

```
{

aoi <- read_sf('Ghana shp file/ROI/new_roi.shp')

aoi <- st_transform(aoi, st_crs(4326))

aoi.ee <- st_bbox(aoi) %>%

st_as_sfc() %>%

sf_as_ee() #Converts it to an Earth Engine Obj

}
```

```
Date <- Sys.Date()
```

Map each image from 2000 to extract the monthly Climatic Data from the Terraclimate dataset and rename the bands of the image

```
{

Precipitation  <- ee$ImageCollection("UCSB-CHG/CHIRPS/DAILY") %>%

ee$ImageCollection$filterDate("2000-01-01", rdate_to_eedate(Date)) %>%

ee$ImageCollection$map(function(x) x$select("precipitation")) %>%

ee$ImageCollection$toBands() # from imagecollection to image


}

{

MinimumTemperature <- ee$ImageCollection("IDAHO_EPSCOR/TERRACLIMATE") %>%

ee$ImageCollection$filterDate("2000-01-01", rdate_to_eedate(Date)) %>%

ee$ImageCollection$map(function(x) x$select("tmmn")) %>%

ee$ImageCollection$toBands()}

{

MaximumTemperature <- ee$ImageCollection("IDAHO_EPSCOR/TERRACLIMATE") %>%

ee$ImageCollection$filterDate("2000-01-01", rdate_to_eedate(Date)) %>%

ee$ImageCollection$map(function(x) x$select("tmmx")) %>%

ee$ImageCollection$toBands()}

{

Evapotranspiration <- ee$ImageCollection("NASA/FLDAS/NOAH01/C/GL/M/V001") %>%

ee$ImageCollection$filterDate("2000-01-01", rdate_to_eedate(Date)) %>%

ee$ImageCollection$map(function(x) x$select("Evap_tavg")) %>%

ee$ImageCollection$toBands()}

{

Humidity <- ee$ImageCollection("NASA/FLDAS/NOAH01/C/GL/M/V001") %>%
```

```
ee$ImageCollection$filterDate("2000-01-01", rdate_to_eedate(Date)) %>%
ee$ImageCollection$map(function(x) x$select("Qair_f_tavg")) %>%
ee$ImageCollection$toBands()}
{
Drought <- ee$ImageCollection("IDAHO_EPSCOR/TERRACLIMATE") %>%
ee$ImageCollection$filterDate("2000-01-01", rdate_to_eedate(Date)) %>%
ee$ImageCollection$map(function(x) x$select("pdsi")) %>%
ee$ImageCollection$toBands()}
```

Extract monthly precipitation values from the Terraclimate ImageCollection through $ee_extract$. $ee_extract$ works similar to $raster :: extract$, you just need to define: the ImageCollection object (x), the geometry (y), and a function to summarize the values (fun).

```
        {
        Precipitation       <- ee_extract(x = Precipitation , y = aoi.ee,
        sf = FALSE,scale = 250, fun = ee$Reducer$mean(), via = "drive",
        quiet = T)

        MinimumTemperature  <- ee_extract(x = MinimumTemperature ,
        y = aoi.ee, sf = FALSE,scale = 250, fun = ee$Reducer$mean(),
         via = "drive", quiet = T)

        MaximumTemperature  <- ee_extract(x = MaximumTemperature ,
        y = aoi.ee, sf = FALSE,scale = 250, fun = ee$Reducer$mean(),
        via = "drive", quiet = T)

        Evapotranspiration  <- ee_extract(x = Evapotranspiration ,
        y = aoi.ee, sf = FALSE,scale = 250, fun = ee$Reducer$mean(),
        via = "drive", quiet = T)
```

```
        Humidity            <- ee_extract(x = Humidity , y = aoi.ee,
        sf = FALSE, scale = 250,fun = ee$Reducer$mean(), via = "drive",
        quiet = T)

        Drought             <- ee_extract(x = Drought , y = aoi.ee,
        sf = FALSE, scale = 250,fun = ee$Reducer$mean(), via = "drive",
         quiet = T)
        }
```

Save the Data to an excell file

```
{
write.csv(Precipitation,"Data/Precipitation.csv")
write.csv(MinimumTemperature,"Data/MinimumTemperature.csv")
write.csv(MaximumTemperature,"Data/MaximumTemperature.csv")
write.csv(Evapotranspiration,"Data/Evapotranspiration.csv")
write.csv(Humidity,"Data/Humidity.csv")
write.csv(Drought,"Data/Drought.csv")
}
```

Tidy the Data

```
{
Precipitation <- Precipitation%>%
pivot_longer(starts_with("X20"),names_to =  c("X","Date"),
names_pattern = "(.)(.+)",values_to = "Precipitation")%>%
separate(Date,into = c("Date","Pr"),sep = "_")%>%
```

```
separate(Date, into = c('year', 'month'), sep = -2, convert = TRUE)%>%
select(year,month,Precipitation)}
{
MinimumTemperature <- MinimumTemperature%>%
pivot_longer(starts_with("X20"),names_to =  c("X","Date"),
names_pattern = "(.)(.+)",values_to = "MinTemperature")%>%
separate(Date,into = c("Date","tmmn"),sep = "_")%>%
separate(Date, into = c('year', 'month'), sep = -2, convert = TRUE)%>%
select(year,month,MinTemperature)}
{MaximumTemperature <- MaximumTemperature%>%
pivot_longer(starts_with("X20"),names_to =  c("X","Date"),
names_pattern = "(.)(.+)",values_to = "MaxTemperature")%>%
separate(Date,into = c("Date","tmmx"),sep = "_")%>%
separate(Date, into = c('year', 'month'), sep = -2, convert = TRUE)%>%
select(year,month,MaxTemperature)}
```

Appendix Chapter 2

```
getQABits <- function(image, qa) {

# Convert binary (character) to decimal (little endian)

qa <- sum(2^(which(rev(unlist(strsplit(as.character(qa), "")) == 1))-1))

# Return a mask band image, giving the qa value.

image$bitwiseAnd(qa)$lt(1)

} mod.clean <- function(img) {

# Extract the NDVI band

ndvi_values <- img$select("NDVI")

# Extract the quality band

ndvi_qa <- img$select("SummaryQA")

# Select pixels to mask

quality_mask <- getQABits(ndvi_qa, "11")

# Mask pixels with value zero.

ndvi_values$updateMask(quality_mask)$divide(ee$Image$constant(10000))

#0.0001 is the MODIS Scale Factor

}

   Date <- Sys.Date()

modis.evi <- ee$ImageCollection("MODIS/006/MOD13Q1")$filter(ee$Filter$

date('2000-01-01',rdate_to_eedate(Date)))$map(mod.clean)


cc.proj <- st_transform(cc, st_crs(2992))

hex <- st_make_grid(x = cc.proj, cellsize = 17080, square = FALSE) %>%

st_sf() %>%
```

```
rowid_to_column('hex_id')
hex <- hex[cc.proj,]
plot(hex)
{
cc.evi <- ee_extract(x = modis.evi, y = hex["hex_id"], sf = FALSE, scale = 250,
 fun = ee$Reducer$mean(), via = "drive", quiet = T)
evi.df <- as.data.frame(cc.evi)
write.csv(x = evi.df, file = "Data/rgeedf.csv")
}


cc.evi = evi.df <-read.csv("Data/rgeedf.csv")
colnames(evi.df) <- c('hex_id', stringr::str_replace_all(substr(colnames
 (evi.df[, 2:ncol(evi.df)]), 2, 11), "_", "-"))
{
evi.hw.lst <- list()
#Create an empty list, this will be used to house the time series
 projections for each cell.
evi.dcmp.lst <- list()
#Create an empty list, this will be used to house the time series
 decomposition for each cell.
evi.df<-evi.df[,-2]
evi.trend <- data.frame(hex_id = evi.df$hex_id, na.cnt = NA, na.cnt.2 = NA,
 trend = NA, p.val = NA, r2 = NA, std.er = NA, trnd.strngth = NA,
  seas.strngth = NA)
#This data frame will hold the trend data
Dates <- data.frame(date = seq(as.Date('2000-01-01'), Date, "month"))
Dates$month <- month(Dates$date)
Dates$year <- year(Dates$date)
i <- 1
```

```
}

tsv <- data.frame(evi = t(evi.df[i, 2:ncol(evi.df)]))

#converting the data to a transposed data frame

colnames(tsv) <- c("evi")


head(tsv) #let's take a look


na.cnt <- length(tsv[is.na(tsv)])

 #We want to get an idea of the number of entries with no EVI value

evi.trend$na.cnt[i] <- na.cnt

td <- tsv %>%

mutate(month = month(as.Date(rownames(tsv))), year = year(as.Date(rownames(tsv)))) %>%

group_by(year, month) %>%

summarise(mean_evi = mean(evi, na.rm = T), .groups = "keep") %>%

as.data.frame()

head(td)


td$date <- as.Date(paste0(td$year, "-", td$month, "-01"))

dx <- Dates[!(Dates$date %in% td$date),]

dx


dx$mean_evi <- NA

tdx <- rbind(td, dx) %>%

arrange(date)

head(tdx)

write.csv(tdx,"Data/NDVI.csv")

na.cnt <- length(tdx[is.na(tdx)])

evi.trend$na.cnt.2[i] <- na.cnt

#add count of na values to dataframe
```

```
rm(td, dx)
 #remove data we're no longer using, this is a good rule of thumb,
 especially when working with larger datasets.
tdx <- ts(data = tdx$mean_evi, start = c(2000, 1), end = c(2022, 01), frequency = 12)
 #convert data to time series.
plot(tdx,pch = 16,
xlab = "Time", ylab = "EVI ", col = "#2E9FDF")

tdx <- if(na.cnt > 0){imputeTS::na_kalman(tdx, model = "auto.arima", smooth = T)}
 else {
tdx
}
plot(tdx,pch = 16, frame = FALSE,
xlab = "Time", ylab = "EVI ", col = "#2E9FDF")

tdx.dcp <- stl(tdx, s.window = 'periodic')
plot(tdx.dcp,pch = 16, frame = FALSE, col = "#2E9FDF")

Tt <- trendcycle(tdx.dcp)
St <- seasonal(tdx.dcp)
Rt <- remainder(tdx.dcp)
plot(Rt)
plot(Tt)
plot(St)

# The Stationary Signal and ACF
plot(Rt,col= "red", main = "Stationary Signal")
acf(Rt, lag.max = length(Rt),
xlab = "lag", ylab = 'ACF', main = '')
```

```
#The Trend Signal anf ACF

plot(Tt,col= "red",main = "Trend Signal")
acf(Tt, lag.max = length(Tt),
xlab = "lag", ylab = "ACF", main = '')

tseries::adf.test(tdx)

tdx.ns <- data.frame(time = c(1:length(tdx)), trend = tdx - tdx.dcp$time.series[,1])
Tt <- trendcycle(tdx.dcp)
St <- seasonal(tdx.dcp)
Rt <- remainder(tdx.dcp)
trend.summ <- summary(lm(formula = trend ~ time, data = tdx.ns)) #tslm
plot(tdx.ns,pch = 16,
xlab = "Time", ylab = "Trend ", col = "#2E9FDF")
abline(a = trend.summ$coefficients[1,1], b = trend.summ$coefficients[2,1], col='red')


evi.trend$trend[i] <- trend.summ$coefficients[2,1]
evi.trend$trnd.strngth[i] <- round(max(0,1 - (var(Rt)/var(Tt + Rt))), 1)
#Trend Strength Calculation
evi.trend$seas.strngth[i] <- round(max(0,1 - (var(Rt)/var(St + Rt))), 1)
#Seasonal Strength Calculation
evi.trend$p.val[i] <- trend.summ$coefficients[2,4]
evi.trend$r2[i] <- trend.summ$r.squared
evi.trend$std.er[i] <- trend.summ$sigma
evi.trend[i,]
```

```
plot(evi.hw <- forecast::hw(y = tdx, h = 12, damped = T))


evi.trend <- read.csv("Data/rgeedf.csv")


for(i in 1:nrow(evi.df)){
tsv <- data.frame(evi = t(evi.df[i, 2:ncol(evi.df)]))
colnames(tsv) <- c("evi")
na.cnt <- length(tsv[is.na(tsv)])
evi.trend$na.cnt[i] <- na.cnt
if(na.cnt < 263){
td <- tsv %>%
mutate(month = month(as.Date(rownames(tsv))), year = year(as.Date(rownames(tsv)))) %>%
group_by(year, month) %>%
summarise(mean_evi = mean(evi, na.rm = T), .groups = "keep") %>%
as.data.frame()
td$date <- as.Date(paste0(td$year, "-", td$month, "-01"))
dx <- Dates[!(Dates$date %in% td$date),]
dx$mean_evi <- NA
tdx <- rbind(td, dx) %>%
arrange(date)
na.cnt <- length(tdx[is.na(tdx)])
evi.trend$na.cnt.2[i] <- na.cnt
rm(td, dx)
tdx <- ts(data = tdx$mean_evi, start = c(2001, 1), end = c(2019, 11), frequency = 12)
tdx <- if(na.cnt > 0){imputeTS::na_kalman(tdx, model = "auto.arima", smooth = T)}
 else {
tdx
}
tdx.dcp <- stl(tdx, s.window = 'periodic')
```

```
evi.dcmp.lst[[i]] <- tdx.dcp

#This will save our decomposition plots

plot(tdx.dcp)

dev.off()

tdx.ns <- data.frame(time = c(1:length(tdx)), trend = tdx - tdx.dcp$time.series[,1])

Tt <- trendcycle(tdx.dcp)

St <- seasonal(tdx.dcp)

Rt <- remainder(tdx.dcp)

trend.summ <- summary(lm(formula = trend ~ time, data = tdx.ns)) #tslm

evi.trend$trend[i] <- trend.summ$coefficients[2,1]

evi.trend$trnd.strngth[i] <- round(max(0,1 - (var(Rt)/var(Tt + Rt))), 1)


evi.trend$seas.strngth[i] <- round(max(0,1 - (var(Rt)/var(St + Rt))), 1)

 #Seasonal Strength Calculation

evi.trend$p.val[i] <- trend.summ$coefficients[2,4]

evi.trend$r2[i] <- trend.summ$r.squared

evi.trend$std.er[i] <- trend.summ$sigma

evi.hw <- forecast::hw(y = tdx, h = 12, damped = T)

evi.hw.lst[[i]] <- evi.hw

# plot(evi.hw)\

# rm(evi.hw, trend.summ, tdx.ns, tdx.dcp, Tt, St, Rt, tdx, na.cnt)

} else {

evi.ts[[i]] <- NA

}

}

head(evi.trend) #Let's take a peak


evi.trend$system.index <- cc.evi[,1]

hex_trend <- hex %>%
```

```
left_join(evi.trend, by = 'hex_id', keep = F) %>%

replace(is.na(.), 0)

hex_trend <- st_transform(hex_trend, st_crs(4326))


**create a Leaflet Web Map!**


library(classInt)

trend_brks <- classIntervals(hex_trend$trend, n=11, style = "fisher")

colorscheme <- RColorBrewer::brewer.pal(n = 11, 'RdYlGn')

palette_sds <- leaflet::colorBin(colorscheme, domain = hex_trend$trend,

bins=trend_brks$brks, na.color = "#ffffff", pretty = T)


pop <- paste0(

"<b> Hex ID: </b>",hex_trend$hex_id,"<br><b>NA Count:</b>",

 hex_trend$na.cnt+hex_trend$na.cnt.2,"<br><b>Trend: </b>",

 format(round(hex_trend$trend, 4), scientific = FALSE),"<br><b> P-Value:</b>",

 round(hex_trend$p.val, 4),"<br><b>R2: </b>",round(hex_trend$r2, 4),

 "<br><b>Std Err: </b>",round(hex_trend$std.er, 4),"<br><b>Trend Strength:</b>",

 round(hex_trend$trnd.strngth, 2),"<br><b>Seasonal Strength:</b>",

  round(hex_trend$seas.strngth, 4),"<br>"

  )

#Here we're creating a popup for our interactive map.


library(leaflet)

library(dplyr)

map <- hex_trend %>%

leaflet() %>%

setView(5.96475,-1.782181, 9) %>%

addProviderTiles("Esri.WorldTopoMap", group = "Topo Map") %>%
```

```
addProviderTiles("Esri.WorldImagery", group = "Imagery",
options = providerTileOptions(opacity = 0.7)) %>%
addPolygons(
fillColor = ~palette_sds(hex_trend$trend),
fillOpacity = hex_trend$trnd.strngth,
opacity = 0.5,
weight = 0.1,
color='white',
group = "Hexbins",
highlightOptions = highlightOptions(
color = "white",
weight = 2,
bringToFront = TRUE),
popup = pop,
popupOptions = popupOptions(
maxHeight = 250,
maxWidth = 250)) %>%
addLegend(
title = "Trend: lm(EVI ~ Month)",
pal = palette_sds,
values = hex_trend$trend,
opacity = 0.7,
labFormat = labelFormat(
digits = 5)) %>%
addLayersControl(
baseGroups = c("Topo Map", "Imagery"),
overlayGroups = c("Hexbins"),
options = layersControlOptions(collapsed = FALSE)) %>%
addScaleBar(position='bottomleft')
```

```
map
```

VAR

Import Data

```
DATA <- read.csv("Data/Data.csv")%>%

dplyr::select(date,mean_evi,Precipitation,Evapotranspiration,

MiniTemperature,MaxTemperature,Humidity,Drought)

head(DATA)

# Check Missing Values


visdat::vis_dat(DATA)

skimr::skim_tee(DATA)

summary(DATA)
```

```
# Correlation

colnames(DATA)<- c("Date","EVI","Precipitation","Evapiration","TempMin","TempMax",

                  "Humidity","Drought")

plot(corr_coef(DATA))
```

```
# Select a set of predictors with minimal multicollinearity

non_collinear_vars(DATA,TempMin,TempMax,Evapiration,Precipitation,Humidity,Drought,

                    max_vif =3)


# Removed correlated variable

DATA <- DATA%>%

dplyr::select(Date,Drought, TempMin, Precipitation, TempMax, Evapiration,EVI)
```

```
colnames(DATA)

head(DATA)
```

Input Missing Values

```
DATA$EVI <- if(is.na(DATA$EVI) > 0){imputeTS::na_kalman(DATA$EVI,

model = "auto.arima",smooth = T)} else {

DATA$EVI

}

head(DATA)
```

Check if Impute worked

```
vis_dat(DATA)

df<- DATA%>%

dplyr::select(-Date)

# 1st differenced data

# df <- as.data.frame(diff(as.matrix(DATA), lag = 1))

#=========================================================

# Summary of ADF test of level variables

#=========================================================


adf.none  <- list(

EVI = ur.df(DATA$EVI, type='none', selectlags = c("AIC")),

Precipitation = ur.df(DATA$Precipitation, type='none', selectlags = c("AIC")),

Evapiration = ur.df(DATA$Evapiration, type='none', selectlags = c("AIC")),

TempMin = ur.df(DATA$TempMin, type='none', selectlags = c("AIC")),

TempMax = ur.df(DATA$TempMax, type='none', selectlags = c("AIC")),
```

```
Drought = ur.df(DATA$Drought, type='none', selectlags = c("AIC"))
)
adf.drift  <- list(
EVI = ur.df(DATA$EVI, type='drift', selectlags = c("AIC")),
Precipitation = ur.df(DATA$Precipitation, type='drift', selectlags = c("AIC")),
Evapiration = ur.df(DATA$Evapiration, type='drift', selectlags = c("AIC")),
TempMin = ur.df(DATA$TempMin, type='drift', selectlags = c("AIC")),
TempMax = ur.df(DATA$TempMax, type='drift', selectlags = c("AIC")),
Drought = ur.df(DATA$Drought, type='drift', selectlags = c("AIC"))
)
adf.trend  <- list(
EVI = ur.df(DATA$EVI, type='trend', selectlags = c("AIC")),
Precipitation = ur.df(DATA$Precipitation, type='trend', selectlags = c("AIC")),
Evapiration = ur.df(DATA$Evapiration, type='trend', selectlags = c("AIC")),
TempMin = ur.df(DATA$TempMin, type='trend', selectlags = c("AIC")),
TempMax = ur.df(DATA$TempMax, type='trend', selectlags = c("AIC")),
Drought = ur.df(DATA$Drought, type='trend', selectlags = c("AIC"))
)
```

```
summary(adf.none$EVI)
summary(adf.none$Precipitation)
summary(adf.none$Evapiration)
summary(adf.none$TempMin)
summary(adf.none$TempMax)
summary(adf.none$Drought)
```

```
summary(adf.trend$EVI)
summary(adf.trend$Precipitation)
```

```
summary(adf.trend$Evapiration)

summary(adf.trend$TempMin)

summary(adf.trend$TempMax)

summary(adf.trend$Drought)
```

```
summary(adf.drift$EVI)

summary(adf.drift$Precipitation)

summary(adf.drift$Evapiration)

summary(adf.drift$TempMin)

summary(adf.drift$TempMin)

summary(adf.drift$Drought)
```

1st differenced data

```
df <- as.data.frame(diff(as.matrix(DATA[,-1]), lag = 1))

df.adf.none  <- list(

EVI = ur.df(df$EVI, type='none', selectlags = c("AIC")),

Precipitation = ur.df(df$Precipitation, type='none', selectlags = c("AIC")),

Evapiration = ur.df(df$Evapiration, type='none', selectlags = c("AIC")),

TempMin = ur.df(df$TempMin, type='none', selectlags = c("AIC")),

TempMax = ur.df(df$TempMax, type='none', selectlags = c("AIC")),

Drought = ur.df(df$Drought, type='none', selectlags = c("AIC"))

)

df.adf.drift  <- list(

EVI = ur.df(df$EVI, type='drift', selectlags = c("AIC")),

Precipitation = ur.df(df$Precipitation, type='drift', selectlags = c("AIC")),

Evapiration = ur.df(df$Evapiration, type='drift', selectlags = c("AIC")),

TempMin = ur.df(df$TempMin, type='drift', selectlags = c("AIC")),

TempMax = ur.df(df$TempMax, type='drift', selectlags = c("AIC")),
```

```
Drought = ur.df(df$Drought, type='drift', selectlags = c("AIC"))
)
df.adf.trend  <- list(
EVI = ur.df(df$EVI, type='trend', selectlags = c("AIC")),
Precipitation = ur.df(df$Precipitation, type='trend', selectlags = c("AIC")),
Evapiration = ur.df(df$Evapiration, type='trend', selectlags = c("AIC")),
TempMin = ur.df(df$TempMin, type='trend', selectlags = c("AIC")),
TempMax = ur.df(df$TempMax, type='trend', selectlags = c("AIC")),
Drought = ur.df(df$Drought, type='trend', selectlags = c("AIC"))
)
```

The ADF result for our variable from the above R code is generated as follows

```
summary(adf.trend$EVI)
summary(df.adf.trend$Precipitation)
summary(df.adf.trend$Evapiration)
summary(df.adf.trend$TempMin)
summary(df.adf.trend$TempMax)
summary(df.adf.trend$Drought)
```

Interpretation of ADF test follow the general-to-specific approach. As such, three regression models are applied sequentially.

```
#=========================================================
# General-to-Specific Investigation
# The case of EVI variable
#=========================================================


print("Level Variable with Trend")
```

```
cbind(t(df.adf.trend$EVI@teststat), df.adf.trend$EVI@cval)


print("Level Variable with Trend")

cbind(t(df.adf.trend$Precipitation@teststat), df.adf.trend$Precipitation@cval)


print("Level Variable with Trend")

cbind(t(df.adf.trend$Evapiration@teststat), df.adf.trend$Evapiration@cval)


print("Level Variable with Trend")

cbind(t(df.adf.trend$TempMin@teststat), df.adf.trend$TempMin@cval)


print("Level Variable with Trend")

cbind(t(df.adf.trend$TempMax@teststat),df.adf.trend$TempMax@cval)


print("Level Variable with Trend")

cbind(t(df.adf.trend$Drought@teststat), df.adf.trend$Drought@cval)


print("1st Diff. Variable with Drift and Trend")

cbind(t(df.adf.trend$EVI@teststat), df.adf.trend$EVI@cval)


print("1st Diff. Variable with Drift")

cbind(t(df.adf.drift$EVI@teststat), df.adf.drift$EVI@cval)


print("1st Diff. Variable with None")

cbind(t(df.adf.none$EVI@teststat), df.adf.none$EVI@cval)
```

Finally, we can conclude that logarithm of real income contains a unit root and can be stationay time series by differencing the first order. Now that this transformed variable contains no unit root, it can be included in VAR or VECM model.

In this process, the alphanumeric names of test statistics are a little confusing but when we refer the above three specifications of regression equations, the meanings of names of test statistics are clear.

```
EVI <- ts(data = df$EVI, start = c(2000, 1), end = c(2022, 01), frequency = 12)
Precipitation <- ts(data = df$Precipitation, start = c(2000, 1), end = c(2022, 01), frequen
Evapiration <- ts(data = df$Evapiration, start = c(2000, 1), end = c(2022, 01), frequency =
TempMin <- ts(data = df$TempMin, start = c(2000, 1), end = c(2022, 01), frequency = 12)
TempMax <- ts(data = df$TempMax, start = c(2000, 1), end = c(2022, 01), frequency = 12)
Drought <- ts(data = df$Drought, start = c(2000, 1), end = c(2022, 01), frequency = 12)
TimeSeries1 <- cbind(EVI,Precipitation,Evapiration,TempMin,TempMax,Drought)
plot(TimeSeries1)
```

Time Series using VAR

```
EVI <- ts(data = DATA$EVI, start = c(2000, 1), end = c(2022, 01), frequency = 12)
Precipitation <- ts(data = DATA$Precipitation, start = c(2000, 1), end = c(2022, 01),
frequency = 12)
Evapiration <- ts(data = DATA$Evapiration, start = c(2000, 1), end = c(2022, 01),
frequency = 12)
TempMin <- ts(data = DATA$TempMin, start = c(2000, 1), end = c(2022, 01), frequency = 12)
TempMax <- ts(data = DATA$TempMax, start = c(2000, 1), end = c(2022, 01), frequency = 12)
Drought <- ts(data = DATA$Drought, start = c(2000, 1), end = c(2022, 01), frequency = 12)
TimeSeries <- cbind(EVI,Precipitation,Evapiration,TempMin,TempMax,Drought)
```

```
library(TSstudio)
ts_plot(EVI)
ts_plot(Evapiration)
ts_plot(Drought)
plot(TimeSeries)
```

Lag Selection

Type of deterministic Regressors to include. We use none because the time series was made stationary using differencing above.

```
LagSelection <-VARselect(TimeSeries,type = "const", lag.max = 12) #highest lag order
LagSelection$selection
kable(t(LagSelection$criteria))%>%kable_styling(latex_options = c("repeat_header"))


LagSelection1 <-VARselect(TimeSeries1,
type = "const",
lag.max = 12) #highest lag order
LagSelection1$selection
kable(t(LagSelection1$criteria))%>%kable_styling(latex_options = c("repeat_header"))
```

Estimating Models

```
# Creating a VAR model with vars
Model <- VAR(TimeSeries,p= 8,lag.max = 12,season = NULL,  exogen = NULL,type = "const")
summary(Model)
```

```
# Creating a VAR model with vars
Model1 <- VAR(TimeSeries1,p= 10,lag.max = 12,season = NULL,  exogen = NULL,type = "const")
summary(Model1)
```

```
predict(Model1, n.ahead = 12, ci = 0.95)
```

forecast is a generic function for forecasting from time series or time series models. The function invokes particular methods which depend on the class of the first argument.

```
forecast(Model1)
plot(forecast(Model1))
```

```
accuracy(forecast(Model1),d=10, D= 1)
```

```
library(vars)
colnames(TimeSeries1) <-c("EVI","Prep","Evap","TMin","TMax","Drght")
Model2 <- VAR(TimeSeries1,p= 10,lag.max = 12,season = NULL,  exogen = NULL,type = "const")
```

```
plot.varfevd  <-function (x, plot.type = c("multiple", "single"), names = NULL,
main = NULL, col = NULL, ylim = NULL, ylab = NULL, xlab = NULL,
legend = NULL, names.arg = NULL, nc, mar = par("mar"), oma = par("oma"),
addbars = 1, ...)
{
K <- length(x)
ynames <- names(x)
plot.type <- match.arg(plot.type)
if (is.null(names)) {
names <- ynames
}
else {names <- as.character(names)
if (!(all(names %in% ynames))) {
warning("\nInvalid variable name(s) supplied, using first variable.\n")
names <- ynames[1]
}
}
nv <- length(names)
#    op <- par(no.readonly = TRUE)
```

```
ifelse(is.null(main), main <- paste("FEVD for", names), main <- rep(main,
nv)[1:nv])
ifelse(is.null(col), col <- gray.colors(K), col <- rep(col,
K)[1:K])
ifelse(is.null(ylab), ylab <- rep("Percentage", nv), ylab <- rep(ylab,
nv)[1:nv])
ifelse(is.null(xlab), xlab <- rep("Horizon", nv), xlab <- rep(xlab,
nv)[1:nv])
ifelse(is.null(ylim), ylim <- c(0, 1), ylim <- ylim)
ifelse(is.null(legend), legend <- ynames, legend <- legend)
if (is.null(names.arg))
names.arg <- c(paste(1:nrow(x[[1]])), rep(NA, addbars))
plotfevd <- function(x, main, col, ylab, xlab, names.arg,
ylim, ...) {
addbars <- as.integer(addbars)
if (addbars > 0) {
hmat <- matrix(0, nrow = K, ncol = addbars)
xvalue <- cbind(t(x), hmat)
barplot(xvalue, main = main, col = col, ylab = ylab,
xlab = xlab, names.arg = names.arg, ylim = ylim,
legend.text = legend, ...)
abline(h = 0)
}
else {xvalue <- t(x)
barplot(xvalue, main = main, col = col, ylab = ylab,
xlab = xlab, names.arg = names.arg, ylim = ylim,
...)
abline(h = 0)
}
```

```
}
if (plot.type == "single") {
for (i in 1:nv) {
plotfevd(x = x[[names[i]]], main = main[i], col = col,
ylab = ylab[i], xlab = xlab[i], names.arg = names.arg,
ylim = ylim, ...)
}
}
else if (plot.type == "multiple") {
if (missing(nc)) {
nc <- ifelse(nv > 4, 2, 1)
}
nr <- ceiling(nv/nc)
par(mfcol = c(nr, nc), mar = mar, oma = oma)
for (i in 1:nv) {
plotfevd(x = x[[names[i]]], main = main[i], col = col,
ylab = ylab[i], xlab = xlab[i], names.arg = names.arg,
ylim = ylim, ...)
}
}
#    on.exit(par(op))
}
```

```
win.graph(width=13,height=8)
layout(matrix(1:6,ncol=1))
plot.varfevd(fevd(Model2, n.ahead = 10 ),plot.type = "multiple", col=1:6)
```

Impulse Responds Analysis

```
plot(irf(Model2,impulse = "EVI",response = "EVI"))

plot(irf(Model2,impulse = "EVI",response = "Prep"))

plot(irf(Model2,impulse = "EVI",response = "Evap"))

plot(irf(Model2,impulse = "EVI",response = "TMin"))

plot(irf(Model2,impulse = "EVI",response = "TMax"))

plot(irf(Model2,impulse = "EVI",response = "Drght"))
```

```
grangertest(EVI~Precipitation, order = 12, data = TimeSeries1)

grangertest(EVI~TempMin, order = 12, data = TimeSeries1)

grangertest(EVI~TempMax, order = 12, data = TimeSeries1)

grangertest(EVI~Evapiration, order = 12, data = TimeSeries1)

grangertest(EVI~Drought, order = 12, data = TimeSeries1)
```

```
plot(forecast(Model1))
```