Doctoral Thesis

Forecasting Vegetation Condition using Remote Sensing Time Series Data

by

Omar Mwana Said

March 2022

Graduate School of Applied Informatics
University of Hyogo

Abstract

Vegetation is a vital terrestrial component and plays an important role in the ecosystem and climate. However, human activities cause changes in vegetation cover leading to vegetation degradation with profound effects on the environment and socio-economic development. Thus, timely monitoring and forecasting vegetation condition is essential for biodiversity and ecosystem conservation. Although vegetation degradation has become a global issue and has negatively impacted many regions, field data crucial for monitoring vegetation condition is rare or non-existent in many regions, notably in east Africa, thus, vegetation indices (VIs) derived from remote sensing data provides useful information and can be applied in monitoring vegetation condition in the region. The normalized difference vegetation index (NDVI) is the one of the most widely used index in vegetation studies which helps to reveal vegetation vigor in regions. The main objective of this research was to develop time series forecasting models using univariate pixel-wise NDVI data between 2000 and 2019 to monitor and forecast vegetation condition in east Africa. Pixel-wise NDVI data was extracted from the Moderate Resolution Imaging Spectroradiometer (MODIS) Terra Vegetation Indices, (MOD13Q1 product, version 6) at 16 - days temporal resolution and 250 m spatial resolution. Holt-Winters, SARIMA and three MLP models MLP 1 (3000 epochs, ReLU AF, Adam optimizer), MLP 2 (4000 epochs, TanH AF, Adamax optimizer) and MLP 3 (5000 epochs, ReLU AF, Nadam optimizer). were developed. The models were used to forecast NDVI in east Africa, in a low (Middle Tana River Basin) and high vegetation (Central and Western Uganda and Ituri province in the Democratic Republic of Congo, DRC) regions. The temporal analysis of NDVI for 1 × 1 pixel in the low and high vegetation regions showed a temporal pattern with distinct low and peak values with the year 2011 marked by a low NDVI in both regions. The observed and predicted NDVI time series graphs for 1 × 1 pixel in the low and high vegetation regions showed that the Holt-Winters and MLP models underestimated as well as overestimated NDVI during most of the study period. On the other hand, the SARIMA model did not provide a good fit leading to poor estimation. Spatial analysis showed that the low vegetation region was forecasted to experience moderate to light vegetation deficit while the high vegetation region was forecasted to have moderate to

extremely high vegetation condition for the 6 months' lead. Moreover, the low vegetation region was predicted to have NDVI between 0.0 to 0.2 during the October – December 2019 and January – March 2020 seasons, while the high vegetation region was predicted to have NDVI between 0.5 to 1.0 during the same seasons. Comparative analysis of the forecasting performance for the models at 6 months lead time were assessed using the root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE) and mean percentage error (MAPE). The Holt-Winters model achieved better performance for 600 × 600 pixels in the high vegetation region, with a RMSE of 0.084, MAE of 0.0679 and MSE of 0.0071, while the MLP 3 model achieved the worst predictive performance. On the other hand, the MLP 2 model had the best predictive performance in the low vegetation region with RMSE of 0.07207, MSE of 0.00589 and MAE of 0.06417, while the SARIMA model was the worst performing model. The results of this thesis are crucial for guiding land use planners and forest managers to formulate mitigation strategies and assist in ecosystem and biodiversity conservation. Moreover, the research may be used as a foundation for future studies using pixel-wise data and for assessing the performance of univariate time series forecasting models.

Table of Contents

Abstract ii
Table of Contents iv
List of Figures vi
List of Tables ix
List of Abbreviations xi
CHAPTER ONE 1
INTRODUCTION 1
1.1 Background information 1
1.2 Problem statement 4
1.3 Research objectives 5
1.4 Research questions 5
1.5 Justification of this study 6
1.6 Scope of the study 6
1.7 Dissertation layout 6
CHAPTER TWO 8
LITERATURE REVIEW 8
2.1 Overview 8
2.2 Remote sensing data 8
2.3 Time series 13
2.4 Time series forecasting models 15
CHAPTER THREE 40
MATERIALS AND METHODOLOGY 40
3.1 Study area 40
3.2 The normalized difference vegetation index data 43
3.3 Virtual machine 46
3.4 Hyperparameters optimization 47
3.5 Implementation of the forecasting models 49
3.6 Models validation metrics 54
CHAPTER FOUR 56

RESULTS AND DISCUSSIONS 56
4.1 Introduction 56
4.2 Temporal analysis using 1 × 1 pixel NDVI data in the low and high vegetation regions 57
4.3 Analysis of results using the Holt-Winters model 60
4.4 Analysis of results using the MLP models 67
4.5 Analysis of results using the SARIMA model 77
4.6 Comparative analysis of the models in forecasting NDVI in the low and high vegetation regions 80
CHAPTER FIVE 82
CONCLUSION 82
ACKNOWLEDGEMENTS 87
REFERENCES 88

List of Figures

Figure 2. 1. An artificial neuron 22
Figure 2. 2. Architectures of forecasting ANN models: (a) Recursive multistep neural network and (b) Direct multistep neural network 27
Figure 2. 3. Architectural graph of MLP with one hidden layer 28
Figure 2. 4. Stages in the development of the Box-Jenkins model 32
Figure 2. 5. Flowchart showing summary of the stages in the construction of a SARIMA model 39
Figure 3.1. Map of the low vegetation region (Middle Tana River Basin) and high vegetation region (Western and Central Uganda and Ituri Province in DRC) (Source: Author). 42
Figure 3.2. Satellite map of the low vegetation region (Middle Tana River Basin) and high vegetation region (Western and Central Uganda and Ituri Province in DRC) (Source: Author; Satellite map source: Google). 43
Figure 4.1 . 1 × 1 pixel NDVI time series in the low vegetation region from February 2000 to September 2019. The pixel was extracted from the top right corner of the h21v09 MODIS tile. 59
Figure 4.2. 1 × 1 pixel NDVI time series in the high vegetation region from February 2000 to September 2019. The pixel was extracted from the bottom left corner of the h21v08 MODIS tile. 60
Figure 4.3. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data using Holt-Winters model (green) in the low vegetation region. 61
Figure 4. 4. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data (green) using Holt-Winters model in the high vegetation region 62
Figure 4. 5. Spatial distribution for the 6 months NDVI forecast using Holt-Winter model in the low vegetation region 64
Figure 4. 6.Spatial distribution for the 6 months NDVI forecast using Holt-Winters model in the high vegetation region. 64

Figure 4. 7. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data (green) using MLP 1 model in the low vegetation region 68
Figure 4. 8. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data (green) using MLP 2 model in the low vegetation region 68
Figure 4. 9. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data (green) using MLP 3 model in the low vegetation region 69
Figure 4. 10. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data (green) using MLP 1 model in the high vegetation region 69
Figure 4. 11. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data (green) using MLP 2 model in the high vegetation region 70
Figure 4. 12. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data (green) using MLP 3 model in the high vegetation region 70
Figure 4. 13. Spatial distribution for the 6 months lead NDVI forecast maps produced by MLP 1 in the low vegetation region. 72
Figure 4. 14. Spatial distribution for the 6 months lead NDVI forecast maps produced by MLP 2 in the low vegetation region 73
Figure 4. 15. Spatial distribution for the 6 months lead NDVI forecast maps produced by MLP 3 in the low vegetation region 73
Figure 4. 16. Spatial distribution for the 6 months lead NDVI forecast maps produced by MLP 1 in the high vegetation region 74
Figure 4. 17. Spatial distribution for the 6 months lead NDVI forecast maps produced by MLP 2 in the high vegetation region 75
Figure 4. 18. Spatial distribution for the 6 months lead NDVI forecast maps produced by MLP 3 in the high vegetation region 75
Figure 4. 19. 1 ×1 pixel monthly mean NDVI time series, showing the training data (blue), testing data (red) and the predicted data using SARIMA model (green) in the low vegetation region. 78

Figure 4. 20. 1 × 1 pixels monthly mean NDVI time series, showing the training data (blue), testing data (red) and the predicted data using SARIMA model (green) in the high vegetation region. 78

List of Tables

Table 2.1. Types of exponential smoothing methods 18
Table 2.2. Summary of model identification using the sample ACF. 34
Table 2.3. Summary for the behaviour of identification functions for SARIMA models.... 38
Table 3. 1. Characteristics of the study area 41
Table 3. 2. Summary of the specifications of of the virtual machine from Oracle Cloud Infrastructure (OCI) 47
Table 3. 3. Search space for hyperparameters for the Holt-Winters model 48
Table 3. 4. Search space for hyperparameters for the MLP model 48
Table 3. 5. Search space for hyperparameters for the SARIMA model 48
Table 3. 6. Summary of the specifications of Subsystem B of the supercomputer from the Research Institute for Information Technology, Kyushu University 52
Table 4.1. Evaluation of the Holt-Winters additive model in the low and high vegetation regions for 10 × 10 pixels. (All computations were done on a remote server, instance Shape: VM.Standard2.8). 66
Table 4. 2. Evaluation of the Holt-Winters additive model in the low and high vegetation regions for 100 × 100 pixels. (All computations were done on a remote server, instance Shape: VM.Standard2.8). 66
Table 4.3. Evaluation of the Holt-Winters additive model in the low and high vegetation regions for 600 × 600 pixels. (All computations were done on a remote server, instance Shape: VM.Standard2.8). 66
Table 4. 4. Forecast error measures between the actual and predicted NDVI for the low and high vegetation regions. 67
Table 4. 5. Evaluation of MLP models in the low vegetation region for 600 × 600 pixels. (All computations were done on the supercomputer) 76
Table 4. 6. Evaluation of MLP models in the high vegetation region for 600 × 600 pixels. (All computations were done on the supercomputer) 76

Table 4. 7. Evaluation of the SARIMA model in the low and high vegetation regions for 100 × 100 pixels. (All computations were done on the supercomputer) 79
Table 4.8. Evaluation of the models using statistical metrics for 600 × 600 pixels in the low vegetation region 81
Table 4. 9. Evaluation of the models using statistical metrics for 600 × 600 pixels in the high vegetation region 81

List of Abbreviations

ACF	Autocorrelation function
AD	Availability Domain
Adam	Adaptive momentum estimation
AF	Activation function
AIC	Akaike information criterion
ANN	Artificial neural networks
AR	Autoregressive
ARIMA	Autoregressive integrate moving average
ASAL	Arid and semi-arid lands
a.s.l.	Above sea level
AVHRR	Advanced Very High Resolution Radiometer
CHIRPS	Climate Hazards Group Infrared Precipitation with Station
CPU	Central processing unit
CSV	Comma separated values
DDR4	Double data rate 4
DMSNN	Direct multistep neural network
DRC	Democratic Republic of Congo
EDR	Enhanced data rate
EOS	Earth observing system
EVI	Enhanced Vegetation Index
FPE	Final prediction error
GB	Gigabytes
Gbps	Gigabytes per second
GHz	Giga Hertz
GPU	Graphic processing unit
GDAL	Geospatial data abstraction library
HBM2	High bandwidth memory 2
HDD	Hard drive disk

HDF	Hierarchical data format
IACF	Inverse autocorrelation function
IOPS	Input/Output Operations Per Second
IPACF	Inverse partial autocorrelation function
LES	Linear exponential smoothing
LiDAR	Light detection and ranging
LP DAAC	Land Processes Distributed Active Archive Center
LST	Land surface temperature
MA	Moving average
MAE	Mean absolute error
MLE	Maximum likelihood
MAPE	Mean absolute percentage error
MODIS	Moderate Resolution Imaging Spectroradiometer
MIR	Middle Infrared
MSE	Mean square error
MTRB	Middle Tana River Basin
Nadam	Nesterov-accelerated adaptive moment estimation
NDVI	Normalized difference vegetation index
NIR	Near Infrared
NOAA	National Oceanic and Atmospheric Administration
OCI	Oracle Cloud Infrastructure
OCPU	Oracle Compute Unit
OS	Operating system
PACF	Partial autocorrelation function
PB	Petabyte
PCI	Peripheral component interconnect
PFLOPS	Peta floating-point operations per second
RADAR	Radio detection and ranging
ReLU	Rectified linear activation

RMSE	Root mean square error
RMSNN	Recursive multistep neural network
SAR	Synthetic aperture radar
SARIMA	Seasonal autoregressive integrate moving average
SARMA	Seasonal autoregressive moving average
SATA	Serial advanced technology attachment
SES	Simple exponential smoothing (single exponential smoothing)
SDSs	Scientific datasets
SGD	Stochastic gradient descent
SIN	Signal to noise
TanH	Hyperbolic tangent
TB	Tera byte
TIR	Thermal infrared
TRMM	Tropical Rainfall Measuring Mission
UL	Upper left
UN SDGs	United Nations Sustainable Development Goals
UAS	Unmanned aerial systems
VI	Vegetation index
VIs	Vegetation indices
VM	Virtual machine
VNICs	Virtual network interface cards

CHAPTER ONE

INTRODUCTION

Chapter one provides information on the motivation for this research and the background information on the research topic. The chapter also describes the problem statement, research questions, objectives, justification and the scope of the research. The chapter ends with the description of the layout of the thesis.

1.1 Background information

Vegetation covers a substantial portion of the earth's land surface (Rodrigues, 2014), plays a key role in the global climate system (Musau et. al., 2018; Du et. al., 2015; Peng, et. al., 2015) and strongly influences animal distributions and dynamics (Pettorelli, et. al. 2006). Dynamic changes in vegetation level are controlled by the balance between climatic and anthropogenic factors (Gamoun, 2013; Bhandari, 2011). However, vegetation has been greatly affected by human activities such as urbanization, clearing of land for cultivation, overgrazing and demand for firewood (Le Houérou, 2009), contributing to the destruction of 1000 million hectares of the world's land (Walling & Fang, 2003). In addition, human activities are placing increasing pressures on land resources creating conflicts (United Nations, 2015) as well as exposing the land surface to erosion, which is exacerbated by poor agricultural practices. The severity of global biodiversity loss and degradation of ecosystems has negative impacts on food security, nutrition, access to water and people's health worldwide (United Nations, 2015). Thus, monitoring and forecasting vegetation condition is important for ecosystem and biodiversity conservation. In addition, accurate and timely information on vegetation condition is crucial for policy makers, forest managers and land use planners.

Rangelands constitute the greatest land area in the world and it refers to region with shrubs and grasslands as the predominate vegetation and may vary from mild sub-arid wooded savannah to desert (FAO, 1986; Pellant et. al., 2020; Rukikaire, 2021). Rangelands are primarily used for livestock production such as in the Sahelian and sub-sahelian zones of

Africa (FAO, 1986). However, they are experiencing high grazing pressure and rapid decline in tree cover (Waswa, 2012). The sub-Saharan Africa region accounts for more than 950 million people (approximately 13% of the global population) where, the agricultural sector plays a critical role in the economy by employing more than half of the total work force (OECD/FAO, 2016).

The east Africa region exhibits diversity in topography, climate and vegetation. The temperature in the region ranges from as low as 12 ºC at the mountain summits and 34 ºC in the lowlands. The lowlands cover a large part of the region and are characterized by a semi-arid climate with an average rainfall of 200 mm whereas the highlands receive above 1500 mm of rainfall (Nicholson, 1996). The region has a considerable range of vegetation types, from tropical forests to semi-dessert grasslands with rangelands covering 79% of the land surface in east Africa (Justice et. al., 1986). The semi-arid regions are characterized by short vegetation cycles that show large interannual variability both in timing and magnitude of vegetation productivity (Cheng, et. al., 2020). Vegetation cover in east Africa is experiencing rapid decline through land conversion for agriculture, unsustainable harvesting of timber and charcoal production as well as poor land management. Moreover, population increase and economic development in the region has accelerated deforestation and degradation (Ahrends et. al., 2010; Pfeifer et. al., 2012; Abera, 2020) leading to a decline in forest and woodland cover by 5.1% and 15.8% respectively between 2002 and 2008 (Pfeifer et. al., 2013). Moreover, vegetation is affected by recurrent severe drought events and agricultural expansion (Abera, 2020), while most forests suffer from rapid deforestation (Waswa, 2012). For instance, the agricultural area in the region increased to 57% between 1975 – 2000 at the expense of natural vegetation, which decreased by 21%, with nearly 5 million hectares of natural vegetation lost per year (Brink & Eva, 2009). In addition, there is insufficient monitoring of vegetation with 21% of the land experiencing a decline in vegetation cover (Waswa, 2012) and a decrease in natural vegetation formations such as forests, bushlands, grasslands and wetlands (Githui et. al., 2009).

Intensive and long-term ground-based assessments can produce fine scale, accurate and localized information, however, they are expensive, time and labor intensive and do not

provide full coverage and visualization of the spatial dynamism in rangeland ecosystems (Ndungu et. al., 2019; Loughran 1989). Furthermore, field data is difficult to use for predicting regional or global changes as such data are traditionally collected at small spatial and temporal scales and vary in their type of reliability (Pettorelli, et. al., 2006). Additionally, the scarcity of in-situ data is a major barrier in vegetation studies in eastern Africa (Abera, 2020), thus, remote sensing imagery has gained wide application in vegetation studies from local to global scales (Rodrigues, 2014). Remote sensing often utilizes ground based sensors or satellite imagery (Hatfield, 2008) and the remote sensing images reveals widespread changes in vegetation across the globe and is one of the main tools for monitoring vegetation condition in a given region (Jamali et. al., 2015).

Vegetation indices (VIs) derived from remote sensing data are useful indicators of vegetation vigor (Campbell, 1987) and may reveal regions with thriving or stressed vegetation. The NDVI is one of the earliest remote sensing analytical product used to simplify the complexities of multi-spectral imagery and the most popular index for vegetation assessment (Huang et. al., 2021). The indicator was developed for studying vegetation traits (Kriegler et. al., 1969; Rouse et. al., 1974) and has the ability to delineate vegetation and vegetative stress (Huang et. al., 2021). The index is an important vegetation measure as it is sufficiently stable to permit meaningful comparisons of seasonal and inter-annual changes in vegetation growth and activity (Huete et. al., 2002). The advantage of the NDVI is in its rationing concept which reduces many forms of multiplicative noise present in multiple bands (Huete, et. al., 2002), mathematical simplicity and ease of interpretation (Julien, 2008).

Time series analysis of remotely sensed data has gained attention in recent times and are indispensable tools in environmental monitoring such as in vegetation studies (Eerens et. al., 2014). Analysis of NDVI time series and NDVI based metrics are widely used to monitor temporal and spatial dynamics of vegetation (Busetto et. al., 2008). However, applications of such methods in forecasting of vegetation indices and dynamic behaviors remains poorly explored (Mutti et. al., 2020). Furthermore, forecasting vegetation is still inadequate in the east Africa region, thus the need to develop vegetation condition forecasting models for the region.

Numerous studies have been carried out using stochastic models and more recently using artificial neural networks (ANN) for monitoring environmental variables (Mutti et. al., 2020; Belayneh et. al., 2016). Stochastic models provide flexibility and systematic approach for the choice of model hyperparameters (Camelo et. al., 2018), and enables to produce a threshold within which the estimates can be tested statistically and used as indications of the accuracy of the model (Institute and Faculty of Actuaries, 1997). On the other hand, the ANN are designed based on the human brain functioning with a variety of mathematical functions to enhance the ability of the models for accurate predictions (Saffariha et. al., 2021). Such forecasting models can be used to estimate vegetation condition both temporally and spatially to aid in decision making and to prepare effective mitigation measures to reduce effects of vegetation degradation at the local and global scale. Thus, in this research, development and assessment of forecasting models for vegetation condition in east Africa using univariate pixel-wise NDVI data was accomplished.

1.2 Problem statement

Vegetation degradation has become a global issue and has negatively impacted the environment and socio-economic development of various regions, notably in east Africa. Vegetation degradation in the region is attributed to natural and human induced factors such as deforestation and overgrazing. Despite the adverse effects of vegetation degradation occurrence in the east Africa region, limited research has been conducted to forecast vegetation condition at the pixel level. While researches have been conducted on forecasting vegetation condition using NDVI data, most of the studies used coarse spatial resolution NDVI data and the non-pixel-wise modeling approach which has reduced robustness (Mutti et. al., 2020). Moreover, the researches mainly focused on arid and semi-arid regions. According to the United Nations (2015), sustainable management of natural resources contributes to economic growth and social progress, which can be achieved by developing and implementing scientific methods for monitoring and assessing land degradation and desertification. Thus, there is a need for a more robust modeling approach that can reveal

potential vegetation deficit regions at the pixel level using moderate spatial resolution NDVI data to forecast vegetation condition in east Africa.

1.3 Research objectives

The main aim of this research was to develop time series forecasting models using univariate pixel-wise NDVI data to forecast vegetation condition at 6 months' lead time in east Africa, to assist policy makers in decision making and formulation of mitigation plans to reduce the impacts of vegetation degradation in the region.

The specific objectives were to:

1) Use a large dataset in order to cover a larger region which may be applied to forecast vegetation condition in other regions with similar characteristics.
2) Monitor the temporal changes in NDVI based on VI data from 2000 to 2019 in east Africa.
3) Evaluate the performance of the forecasting models in forecasting NDVI at 6 months lead time.
4) Create spatial maps for the 6 months lead NDVI forecasts at the pixel level for easier visualization and identification of potential vegetation deficit regions.

1.4 Research questions

This research aims to answer the following research questions:

1) Can pixel-wise modelling approach using MODIS NDVI data at 250 m spatial resolution provide a robust method to forecast NDVI in east Africa?
2) Which forecasting model is suitable for forecasting NDVI in a low and high vegetation regions in east Africa?
3) How do the developed forecasting models perform in forecasting NDVI in the low and high vegetation regions?
4) Can the spatial maps reveal regions with thriving or deficit vegetation?

1.5 Justification of this study

This work is important for attaining African Union Agenda 2063 goal 7 and its priority areas of environmental sustainability, biodiversity conservation and natural resource management as well as the United Nations Sustainable Development Goals (UN SDGs) goal 15, by providing a methodology required to protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests and combat desertification, halt and reverse land degradation and biodiversity loss (United Nations, 2015). To achieve these goals, mitigation measures to reduce the impacts of vegetation degradation is paramount by using the readily and freely available remote sensing data to detect changes in vegetation condition and develop optimal approaches using forecasting models.

1.6 Scope of the study

Although the east Africa region has a total land area of 5.9 million km^2, the current research focused only on a section of the Middle Tana River Basin in Kenya (henceforth, referred as the low vegetation region) and sections of central and western Uganda and Ituri province in the Democratic Republic of Congo (henceforth, referred as the high vegetation region). The area for the study region was calculated as follows: 1 MODIS tile = 4800 $\times$ 4800 pixels and 1 pixel = 250 m. Therefore, the area for 1 $\times$ 1 pixel (in km^2) = 250 $\times$ 250 = 0.0625 km^2. Thus, each region in the study area comprised of 600 $\times$ 600 pixels, which is equal to an area of 22, 500 km^2. Forecasting the vegetation condition was based on monthly mean pixel-wise NDVI data extracted from the MODIS MOD13Q1 product, at 250 m spatial resolution, from February 2000 to September 2019. The findings of this research can be generalized for other regions in east Africa with similar vegetation condition and biophysical characteristics such as east Ethiopia and west Tanzania (having similarity with the low and high vegetation regions in this research respectively).

1.7 Dissertation layout

This dissertation consists of five chapters. Chapter one presents a general introduction, problem statement, research objectives, research questions, justification and scope of the

research. Chapter two reviews related literature on monitoring and forecasting NDVI and the methodologies used in previous studies. Chapter three describes the study area, data used, methodology, models' implementation and validation metrics. The research findings are presented in Chapter four under results and discussions, while the final chapter (five) provides the conclusion, recommendation and future work.

CHAPTER TWO

LITERATURE REVIEW

Chapter two provides an overview of remote sensing data, types of remote sensors and vegetation indices. This chapter explains time series and its components and different time series forecasting methods such as stochastic and artificial neural networks (ANN) models used in monitoring and forecasting vegetation condition. It also highlights the advantages and disadvantages of the models.

2.1 Overview

Vegetation is an important component of terrestrial ecosystem and a crucial element in the global climate system (Foley et. al., 2000). Studies involved in vegetation dynamics have been active research topics since the late 19th century (Rodrigues, 2014). Incremental, cyclic and periodic changes in vegetation structure and condition are complex and continuous phenomena affected by multiple factors (Bhandari, 2011) such as climate change and human induced factors through extensive land use.

2.2 Remote sensing data

Remote sensing refers to the science of obtaining information about objects or areas from a distance using satellites (NOAA, 2021). It relies on capturing and interpreting the electromagnetic radiation from the earth's surface or objects in space using remote sensors. The sensors collect data by detecting the energy that is reflected from the earth and the sensors can be on satellites or mounted on aircraft (NOAA, 2021).

Remote sensing has become a vital tool and a valuable source of data as it provides regular, open and objective data (Love, 2015) at a global scale with long term records and high spatial resolutions. There are various applications of remote sensing data such as, in monitoring changes in shorelines, measure ocean temperature and wave heights, assessment of natural disasters such as flooding, and natural resource management such as monitoring land use and mapping wetlands (NOAA, 2021).

Moreover, remote sensing has changed the way land resources are viewed, used and managed (Huang, et. al., 2021) and has provided opportunities to monitor and evaluate vegetation condition using spectral reflectance through the use of vegetation indices (VIs). Numerous VIs derived from ratios of spectral reflectance's have been developed, with the NDVI been the most commonly used vegetation index (VI) due to its ease of use, high temporal cycle and global coverage.

2.2.1 Types of remote sensors

As mentioned above, remote sensors capture and interpret electromagnetic radiation from the earth's surface and there are two main types of remote sensors classified according to the source of signal they use to explore objects, namely: active and passive. Active sensors operate with their own source of emission or light while passive sensors rely on reflected light (EOS, 2020; NOAA, 2021; Schoen-Phelan, 2020).

2.2.1.1 Active sensors

Active sensors use internal stimuli to collect data about the earth, i.e., they direct their signals to the object and then checks the response (EOS, 2020; NOAA, 2021; Schoen-Phelan, 2020). Active remote sensing techniques differ by what they transmit (light or waves) and what they determine, such as distance, height, atmospheric condition, e.t.c. (EOS, 2020). The advantage of active sensors is that they fully function at any time of the day as they do not require sunlight and they are relatively dependent of atmospheric scatterings (EOS, 2020).

Examples of active sensors include: Radio detection and ranging (RADAR) which assists in ranging radio signals, light detection and ranging(LiDAR) which determines distance with light, laser altimeter which measures elevation with LiDAR and scatterometer which measures bounced radiation (EOS, 2020).

2.2.1.2 Passive sensors

Passive sensors respond to external stimuli and focus on collecting data from the thermal emissions that are reflected by the earth as the sun's rays hit (Schoen-Phelan, 2020). The

disadvantage of passive sensors is that they depend on sunrays bounced by the target, thus, can only be applied with proper sunlight. Passive sensors employ multispectral or hyperspectral sensors that measure the acquired quantity with multiple band combinations, which differ by the number of channels and includes spectra such as visible, microwave, IR, NIR and thermal infrared (TIR) (EOS, 2020).

Examples of passive sensors include: spectrometer which distinguish and analyze spectral bands, radiometer which determines the power of radiation emitted by the object in particular band ranges such as visible, IR and microwave, spectroradiometer which finds out the power of radiation in several band ranges, hyperspectral radiometer operates with the most accurate type of passive sensor which differentiates hundreds of ultimately narrow spectral bands within visible, NIR and MIR regions, imaging radiometer which scans the object or a surface to reproduce the image, sounder which senses the atmospheric conditions vertically and accelerometer that detects changes in speed per unit of time such as linear or rotational (EOS, 2020).

2.2.2 Quality of remote sensing data

The quality of remote sensing data is considered based on attributes, namely: spatial resolution, which determines the smallest feature that can be detected, sensor footprint, which determines the amount of distortion observed in the data being collected, sampling rate, which determines how much overlap there is between samples, dwell rate, which has an effect on how much time the instrument has to empty old data, spectral resolution, which refers to the indicator of the device's ability to detect very fine wavelengths intervals and is influenced by the atmosphere as some energy can be absorbed by it and signal to noise (SIN) ratio which refers to a measure that compares the level of a desired signal to the level of background noise (Schoen-Phelan, 2020).

2.2.3 Vegetation Indices

Sensors have been developed to measure spectral reflectance at various wavelengths and have been used to monitor vegetation cover, such as vegetation indices (VIs), which have

allowed researchers to relate differences in reflectance to changes in canopy characteristics (Hatfield, et. al., 2008). Spectral VIs utilize the reflectance properties of vegetation and have been widely used to infer temporal and spatial variations in vegetation structure. In addition, the VIs allow reliable spatial and temporal inter-comparisons of terrestrial photosynthetic activity and canopy structural variations (Huete et. al., 2002). Most VIs combine reflectance in the red and near infra-red (NIR) and are generally computed for all pixels temporally and spatially regardless of the biome type, land cover condition and soil type, thus, representing true surface measurements (Rodrigues, 2014). VIs have been applied in various studies and have become vital tools to assess changes in ecosystem with respect to vegetation dynamics and have been particularly helpful in regions with limited in-situ data. The NDVI is the most widely used spectral VIs used to assess live green vegetation in a target (Gitelson, 2012).

The most common method of calculating NDVI is shown in equation (2.1) (Goward et. al., 1991). The formula is based on the fact that chlorophyll absorbs red light while the mesophyll leaf structure scatters near the infrared (Stige, et. al., 2006).

$$NDVI = \frac{NIR-R}{NIR+R} \qquad (2.1)$$

where NIR and R refer to the spectral reflectance in the near infrared and red bands respectively.

Generally, NDVI values range between -1 to +1, where negative values represent water bodies, values close to zero represent rocks, sand or concrete surfaces and positive values represents vegetation such as shrubs, crops, forests etc. Greater NDVI values indicates vigorous vegetation greenness (Jones & Vaughan, 2010), due to light absorption by chlorophyll, thus, the red spectral reflectance is always lower than the NIR for green vegetation (Rodrigues, 2014).

NDVI has been widely used in various studies such as forecasting forage supply (Alhamad et. al., 2007), drought assessment (Sruthi & Aslam, 2015; Karnieli et. al., 2010), analysis of vegetation cover (Sun et. al., 2015), land degradation assessment (Yengoh et. al., 2015), vegetation monitoring (Ahmed, 2016; Anyamba & Tucker, 2005), prediction of African buffalo forage quality (Ryan et. al., 2012), monitoring rangelands (Ndungu et. al.,

2019), animal ecology (Pettorelli et. al., 2011), epidemiological studies (Gascon et. al., 2016) and monitoring vegetation changes in ASALs (Trodd & Dougill, 1998).

The Moderate Resolution Imaging Spectroradiometer (MODIS) launched on the Earth Observing System (EOS) satellites on the EOS-AM platform on June 1998 and the EOS-PM platform in December 2000. MODIS is a 36 channel radiometer covering 0.415 – 14.235 μm wavelengths, with a spatial resolution from 250 m to 1 km at nadir (Running, et. al., 1994). MODIS is designed to provide data on terrestrial biospheric dynamics and process activity (Running, et. al., 1994) and are available in 3 projections: Sinusoidal, Lambert Azimuthal Equal-Area and Geographic, with the Sinusoidal Tile Grid used for most of the higher resolution MODIS land products. Global data from MODIS are organized as units that are 10 deg by 10 deg at the Equator, known as tiles. The tile coordinate system starts at (0, 0) in the upper left (UL) corner and proceeds right (horizontal) and downward (vertical) (Didan, et. al., 2015).

The MODIS vegetation indices (VI) produced at 16 – day intervals provide consistent spatial and temporal comparisons of global vegetation conditions and can be used to monitor the earth's terrestrial photosynthetic vegetation activity in support of phenologic, change detection and biophysical interpretations (Didan et. al., 2015). The MODIS VI algorithm operates on a per pixel basis and applies a filter to the data such that cloud contaminated pixels and extreme off-nadir sensor views are considered lower quality while a cloud free, nadir view pixel with no residual atmospheric contamination represents the best quality pixel, which are used for further compositing (Didan, et. al., 2015). Two VI: normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI) are derived from atmospherically corrected reflectance in the red, near infrared and blue bands (Didan, 2015). The NDVI is the continuity index to the National Oceanic and Atmospheric Administration Advanced Very High Resolution Radiometer (NOAA AVHRR). NDVI time series record for historical and climate applications, while EVI minimizes canopy – soil variations and improves sensitivity over dense vegetation conditions (Didan, 2015). The two VIs complement each other in global vegetation studies and improve upon the detection of vegetation changes and extraction of canopy biophysical parameters (Huete, et. al., 2002).

MODIS NDVI data was selected in this research due to its correlation with many ecosystem attributes, availability of the data and coverage over the study area at 250 m spatial resolution and 16 days' temporal resolution.

2.3 Time series

A series refers to an ordered sequence of observations, thus a time series is a sequence of observations taken sequentially in time (Box & Jenkins, 1970), i.e., it is a set of time-ordered observations of a process where the intervals between the observations remain constant, for example, days, weeks, months and years (Cowpertwait & Metcalfe, 2009). The main feature of a time series is its temporal dependence, i.e., the distribution of an observation at a certain time point is conditional on the previous value of the series and depends on the outcome of those previous observations (Charlton & Caimo, 2012). Therefore, the main difference between time series models and traditional models is that, in time series models, lag values of the target variable are used as predictor variables whereas in traditional models, other variables are used as predictors (Charlton & Caimo, 2012). Time series forecasting is an important area in machine learning as it adds a time dimension in a machine learning dataset and uses the information in a time series to forecast future values of the series. Time series data is distinguished from other types of longitudinal data by the number of source of the observations (Jebb et. al., 2015). Time series data provides a powerful tool to learn from past events, monitor current conditions and prepare for future (van Leeuwen et. al., 2004; Jebb et. al., 2015).

Time series analysis involves developing models that best capture an observed time series in order to understand underlying causes by making assumptions about the form of the data and decomposing the time series into constituent components (Brownlee, 2020). The main aims of a time series analysis are to describe all available data and provide meaningful interpretation to the problem, i.e., to model the stochastic mechanisms that gives rise to an observed series and to forecast the future values of a series based on the history of that series (Cryer & Chan, 2008; Charlton & Caimo, 2012).

Time series analysis of remote sensing data adds a temporal dimension to the data which can be used to examine change, phenology and noise data (Sakamoto et. al., 2005). Remote sensing VIs time series data provide crucial information to track vegetation phenological cycle, thus, analysis of VIs time series enables the extraction of appropriate metrics to describe vegetation condition and dynamics to help us monitor vegetation cover as well as understand vegetation phenological changes (Bradley & Mustard, 2008). Using high temporal resolution remote sensing data such as the ones obtained through the MODIS sensor have been found to contain more information than high spectral or spatial resolution remote sensing data (Mutti et. al., 2020; Key et. al., 2001). Furthermore, consistent NDVI time series data are paramount in monitoring ecological resources that are being altered by climate and human impacts (van Leeuwen et. al., 2006).

2.3.1. Components of time series

The four components of time series are trend, cycle, seasonality and irregular variation, which are described in the sections below.

2.3.1.1. Trend

Trend refers to the systematic change in the level of a series, where both the direction and slope of a trend may remain constant or change throughout the series. The trend in the data is important as it shows variability in the data, thus must be taken into consideration when performing time series analysis, by either modeling it explicitly or removing it from mathematical transformations (McCleary et. al., 1980). The first approach is done when the trend is theoretically interesting while the second approach is done when the trend is not pertinent to the goals of the analysis (Jebb et. al., 2015).

2.3.1.2. Cycle

A cycle refers to a pattern of fluctuation that occurs across a period of time, with the patterns not having a fixed duration. The cycle is a non-seasonal component in time series data, and

may be present with a trend and/or seasonality or without any of these components (Jebb et. al., 2015).

2.3.1.3. Seasonality

The seasonality component of a time series is the repeating pattern of movement within a period of one year or less and is attributed to seasonal factors (Cowpertwait & Metcalfe, 2009). Although the underlying pattern remains fixed, the magnitude of a seasonal effect may vary across periods and can be embedded within overarching trends (Jebb et. al., 2015). Seasonality is a seasonal component of time series data and is associated with calendar feature such as days, weeks or months.

2.3.1.4. Irregular variation (randomness)

The irregular component represents statistical noise and is analogous to the error terms included in various types of statistical models (Jebb et. al., 2015). It constitutes any remaining variation in a time series after the trend, cycle and seasonality components have been partitioned out (Jebb et.al., 2015). The irregular component is referred to as white noise when the component is completely random, and is vital in judging the adequacy of a time series model (Jebb et. al., 2015).

2.4 Time series forecasting models

Forecasting involves making future projections on the basis of historical and current data (Kalekar, 2004). Forecasting models are either deterministic or stochastic. A deterministic model only makes a single set of assumptions about the values of important variables to represent future observations. On the other hand, a stochastic model is one in which at least one significant variable is assumed to vary in a stochastic manner. In such models, the future path of these variables is treated as uncertain and they depend on the random behavior of these variables. Such models allow validity of the assumptions to be tested statistically and produce estimates not only of the expected value of the future observations but also of the variation about the expected value (Institute and Faculty of Actuaries, 1997; Schobel, 2012).

As mentioned in chapter 1, section 1.1, there are various techniques used in vegetation studies such as stochastic and ANN models. Stochastic models commonly applied in forecasting economic indices such as Holt-Winters and SARIMA models have been increasingly used for monitoring ecological and environmental variables (Mutti et. al., 2020).

Fernàndez-Manso et. al., (2011) used SARIMA model to forecast short-term NDVI in each 10-day period using maximum value composite (MVC) bands of the NDVI obtained from NOAA AVHRR data and concluded that time series models could be used for vegetation monitoring. Alhamad et. al. (2006) applied a point based biophysical simulation of forage production coupled with 1 km AVHRR NDVI data to determine the feasibility of projecting forage conditions 84 days into the future using the Box-Jenkins model. They noted that vegetation revealed a bimodal growth pattern which exhibited two periodic peaks of 13 and 26 weeks. They concluded that the ARIMA model offered a promising approach to near-term projection of forage conditions up to 84 days. Mutti et. al., (2020) combined the use of SARIMA, Holt-Winters, Box-Jenkins and SARIMA with an explanatory variable (SARIMAX) and BJT models to model NDVI time series at six desertification hot-spots using MODIS MOD13A2 product at 1 km spatial resolution. They concluded that the use of non-pixel-wise models losses the robustness of pixel-wise models and that NDVI can be forecasted using its past values alone without the need of other explanatory data. Recent studies have applied ANN in environmental monitoring. e Silva et. al., (2020) analyzed changes in land cover to estimate a future scenario for 2035 using ANN for the Taperoà River Basin in northeastern Brazil. They applied the multi-layer perceptron (MLP) model using Landsat satellite images and concluded that the MLP model showed satisfactory results.

Prediction of pixel-wise NDVI in this study was achieved using the Holt-Winters, SARIMA and MLP models which are explained in the sections below.

2.4.1. Exponential smoothing

Exponential smoothing was proposed in the 1950s (Brown, 1959; Holt, 1957; Winters, 1960). It originated with Robert G. Brown in 1944 while working for the US Navy as an Operations Research analyst by developing a tracking model for fire control information on the location

of submarines, which was essentially the simple exponential smoothing of continuous data (Gardner, 2005). Exponential smoothing is the procedure of continually revising a forecast in light of more recent events by assigning exponentially decreasing weights as the observation gets older, without the necessity of fitting a parametric model (Brown, 1959; Kalekar, 2004; Hyndman & Athanasopoulos, 2018; Gelper, et. al., 2010; Hyndman et. al., 2008).

In exponential smoothing, the trend component is in itself a combination of level (l) and a growth term (b), where, the level and growth can be combined in a number of ways, giving five future trend types, as shown in equations (2.2) to (2.6) (Hyndman et. al., 2008). For instance, let T_h denote the forecast trend over the next h time periods, and let ϕ denote a damping parameter ($0 < \phi < 1$). Then, the five trend types are shown in equations (2.2) to (2.6) (Hyndman et. al., 2008).

None: $$T_h = l \tag{2. 2}$$

Additive: $$T_h = l + bh \tag{2. 3}$$

Additive damped: $$T_h = l + (\phi + \phi^2 + \cdots + \phi^h)b \tag{2. 4}$$

Multiplicative: $$T_h = lb^h \tag{2. 5}$$

Multiplicative damped: $$T_h = lb^{(\phi+\phi^2+\cdots+\phi^h)} \tag{2. 6}$$

The damped trend method is appropriate when there is a trend in the time series but the growth rate at the end of the historical data is unlikely to continue more than a short time into the future. The damped equations dampen the trend as the length of the forecast horizon increases in ordered to improve the accuracy at long lead times. After choosing a trend component, a seasonal component is introduced which can be either additive or multiplicative and finally an error is included, either additively or multiplicatively. However, the nature of the error component is often ignored because the distinction between additive and multiplicative errors makes no difference to point forecasts (Hyndman et. al., 2008).

Table 2.1 shows a summary of different types of exponential smoothing methods. Cell NN describes the SES, cell AN describes the Holt-linear method, while the Holt-Winters additive method is given by cell AA and the Holt-Winters multiplicative method is given in cell AM, whereas the other cells corresponds to less commonly used methods (Hyndman et. al., 2008).

Table 2.1. Types of exponential smoothing methods

Trend component	Seasonal component		
	N (none)	A (additive)	M (multiplicative)
N (none)	NN	NA	NM
A (additive)	AN	AA	AM
M (multiplicative)	MN	MA	MM
D (damped)	DN	DA	DM

The three main exponential smoothing methods are: single (simple) exponential smoothing (SES), double exponential smoothing and triple exponential smoothing (Holt-Winters). The SES is applied to data which does not contain trend or seasonal components while the double exponential smoothing model is applied to data which has a trend component. On the other hand, the triple exponential smoothing method (also known as the Holt-Winters model) allows modeling time series' level, trend and seasonality by the use of three updating equations and multi-step forecasting with a simple linear combination of these elements (Brito et.al., 2021). It is worth mentioning that the exponential smoothing models are non-stationary by definition, are able to capture the dynamics of time series and are useful for time series data with non-linear characteristics (Hyndman & Khandakar, 2008).

The most commonly used exponential smoothing models are further described below.

2.4.1.1. Single exponential smoothing (SES) model

The single exponential smoothing (also known as simple exponential smoothing) is the simplest form of the exponential smoothing models, suitable for forecasting data with no

clear trend or seasonal pattern. In SES, forecasts are calculated using weighted averages, where the weights decrease exponentially as observations come from further in the past, and the smallest weights are associated with the oldest observations (Hyndman & Athanasopoulos, 2018). The SES model is expressed as shown in equation (2.7) (Nau, 2020).

$$L_t = \alpha Y_t + (1 - \alpha) L_{t\text{-}1} \qquad (2.7)$$

Where α denotes a smoothing constant, L represent the current level of the series estimated from the data up to the present.

The forecast for the next period is simply the current smoothed value as shown in equation (2.8).

$$\hat{Y}_{t\text{-}1} = L_t \qquad (2.8)$$

2.4.1.2. Double exponential smoothing model

There are two variations of the double exponential smoothing method, namely Brown's method (one smoothing parameter linear method) and Holt's method (two smoothing parameters linear method) (Muchayan, 2019).

Brown's double exponential smoothing, also known as Brown's linear exponential smoothing (LES) uses two different smoothed series that are centered at different points in time (Nau, 2020). The standard form of Brown's LES model is expressed as equation (2.9) (Nau, 2020).

$$S'_t = \alpha Y_t + (1 - \alpha) S'_{t\text{-}1} \qquad (2.9)$$

The forecast Y_{t+k} for any k>1 is expressed as shown in equation (2.10)

$$\hat{Y}_{t+k} = L_t + kT_t \qquad (2.10)$$

where

$$L_t = 2S'_t - S''_{t\text{-}1} \qquad (2.11)$$

$$T_t = (\frac{\alpha}{1-\alpha})(S'_t - S''_{t\text{-}1}) \qquad (2.\ 12)$$

Where, L_t is the estimated level at period t and T_t is the estimated trend at period t

Brown's LES uses a single smoothing parameter which places a constraint on the data patterns as the level and trend are not allowed to vary at independent rates, thus, Holt's LES model addresses the issue by including two smoothing parameters: α and β for the level and trend respectively (Nau, 2020). Holt's LES model is computed using equations (2.13) to (2.15) (Nau, 2020), with the estimate of the level and trend computed as in equation (2.13) and (2.14) respectively, while the forecast is calculated as shown in equation (2.15).

$$L_t = \alpha Y_t + (1-\alpha)(L_{t\text{-}1} + T_{t\text{-}1}) \qquad (2.\ 13)$$

$$T_t = \beta(L_t - L_{t\text{-}1}) + (1-\beta)T_{t\text{-}1} \qquad (2.\ 14)$$

$$\hat{Y}_{t+k} = L_t + kT_t \qquad (2.\ 15)$$

2.4.1.3. Triple exponential smoothing model

The Holt-Winters method was proposed by Holt (1957) and Winters (1960) and uses exponentially weighted moving averages to update estimates of the seasonally adjusted mean (level), slope and seasonals (Cowpertwait & Metcalfe, 2009). The Holt-Winters model is applied on time series data which exhibit patterns of increasing or decreasing trend with the presence of seasonality (Aryee, et. al., 2019; Chatfield, 1978). The model has low dependence on historical data and considers all previous values while giving weight to the most recent values. Moreover, the model requires low data storage, is simple and easy to automate as well as implement and provides better forecasting results compared to other time series models such as ARIMA.

The Holt-Winters model is defined using three smoothing parameters: α, β and γ which are associated with the level, trend and seasonal components respectively. Two variations exist for the Holt-Winters model: the additive and multiplicative models. In the additive case, the time series shows steady seasonal fluctuations regardless of the overall

level of the series while in the multiplicative case, the size of the seasonal fluctuations vary depending on the overall level of the series (Kalekar, 2004; Aryee, et. al., 2019).

The Holt-Winters model with additive seasonals is computed as shown in equations (2.16) to (2.19). The forecasting equation for x_{n+k} made after the observation at time n is shown in equation (2.19) (Cowpertwait & Metcalfe, 2009).

$$a_t = \alpha(x_t - s_{t\text{-}p}) + (1-\alpha)(a_{t\text{-}1} - b_{t\text{-}1}) \qquad (2.16)$$

$$b_t = \beta(a_t - a_{t\text{-}1}) + (1-\beta)b_{t\text{-}1} \qquad (2.17)$$

$$s_t = \gamma(x_t - a_t) + (1-\gamma)s_{t\text{-}p} \qquad (2.18)$$

$$\hat{x}_{n+k|n} = a_n + kb_n + s_{n+k-p} \qquad k \leq p \qquad (2.19)$$

where: a_t, b_t and s_t are the estimated level, trend and seasonal effect respectively at time t, p is the period, and α, β and γ are the smoothing parameters for the level, trend and seasonal components respectively.

The Holt-Winters model with multiplicative seasonal variation is derived as shown in Equations (2.20) to equation (2.23) (Cowpertwait & Metcalfe, 2009).

$$a_n = \alpha\left(\frac{x_n}{s_{n\text{-}p}}\right) + (1-\alpha)(a_{n\text{-}1} - b_{n\text{-}1}) \qquad (2.20)$$

$$b_n = \beta(a_n - a_{n\text{-}1}) + (1-\beta)b_{n\text{-}1} \qquad (2.21)$$

$$s_n = \gamma\left(\frac{x_n}{a_n}\right) + (1-\gamma)s_{n\text{-}p} \qquad (2.22)$$

$$\hat{x}_{n+k|n} = (a_n + kb_n)\, s_{n+k-p} \qquad k \leq p \qquad (2.23)$$

2.4.2. Artificial neural networks (ANN)

Artificial neural networks are supervised learning non -linear models based on the operation of biological neurons (Ornella, et. al., 2020; Abraham, 2005; Mishra & Desai, 2006). They are viable computational models comprised of densely interconnected adaptive processing

units (Hassoun, 1995; Balkin & Ord, 2000) with a wide application in pattern classification, speech synthesis and recognition, adaptive interfaces between humans and complex physical systems, forecasting and prediction, image data compression, etc. (Hassoun, 1995; Graupe, 2013). ANN learn based on a black box process and their performance is affected by input adequacy, network architecture and model validation (Fung, et. al., 2020).

The basic processing elements of neural networks are artificial neurons or nodes, where, each node performs simple computations and each conveys a signal from one node to another, labeled by a weight indicating the extent to which a signal is amplified or diminished by a connection (Mehrothra, et. al., 1997). Figure 2.1 shows a typical artificial neuron, where the signal flow from inputs x_1, ... x_n is considered unidirectional (marked by arrows) as well as the output signal flow (O) (Abraham, 2005).

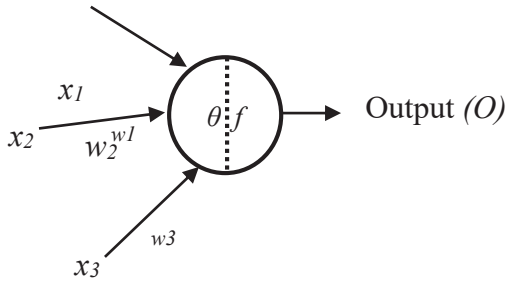

Figure 2. 1. An artificial neuron

The neuron output signal (O) is expressed using the relationship indicated in equation (2.24) (Abraham, 2005).

$$O = f(net) = f\left(\sum_{j=1}^{n} w_i x_j\right) \qquad (2.24)$$

Where, w_j is the weight vector, and the function *f(net)* is the activation (transfer) function. The variable *net* is defined as a scalar product of the weight and input vectors as illustrated in equation (2.25)

$$net = w^{\mathrm{T}} x = w_1 x_1 + \cdots + w_n x_n \qquad (2.25)$$

where T is the transpose of a matrix.

The output value O may also be computed by equation (2.26).

$$O = f\ (net) = \begin{cases} 1 \\ 0 \end{cases} if\ w^T x\ \geq\ \theta\ otherwise \qquad (2.\ 26)$$

Where θ is the threshold level.

2.4.2.1. Neural networks optimizers

Neural networks are difficult and slow to train, thus, optimization methods are used to reduce this difficulty (Riedmiller & Braun, 1992; Kingma & Ba, 2015). There are various optimizers used in neural networks such as Stochastic gradient descent (SGD), Adaptive momentum estimation (Adam), Adamax, Nesterov-accelerated adaptive moment estimation (Nadam) etc. Gradient descent is a simple, robust optimization algorithm that follows the negative gradient of an objective function in order to locate the minimum function (Brownlee, 2021). Adam proposed by Kingma & Ba (2015), is a commonly used optimizer which implements heuristics to estimate the mean and variance of the gradient and are used to generate more stable updates during training (Bello et. al., 2017), while Adamax is a variant of Adam based on the infinity norm (Kingma & Ba 2015). On the other hand, Nadam (Dozat, 2016) is an extension of the Adam algorithm that incorporates Nesterov momentum and can result in better performance of the optimization algorithm (Brownlee, 2021).

2.4.2.2. Back propagation

An ANN is usually learned by adjusting the weights and biases in order to minimize a cost function using the error back propagation algorithm (Ornella et. al., 2020). Back propagation is the popular method for training multi-layer perceptron (MLP) with the term popularized through the publication of the book 'Parallel Distributed Processing' by Rumelhart et. al., (1986). The algorithm represented a landmark in neural networks in that it provided a computationally efficient method for the training of MLP with the training process proceeding in two phases (Haykin, 2009):

1. In the forward phase, the synaptic weights of the network are fixed and the input signal is propagated through the network, layer by layer until it reaches the output,

thus, changes are confined to the activation potentials and outputs of the neurons in the networks.

2. In the backward phase, an error signal is produced by comparing the output of the network with a desired response and the resulting error signal is propagated through the network, layer by layer, however, the propagation is performed in the backward direction. In this phase, successive adjustments are made to the synaptic weights of the networks while the calculation of the adjustments for the output layer is straightforward.

Further details on back propagation algorithm can be found in Rumelhart et. al., (1986) and Bishop (1995).

2.4.2.3. The ANN architecture

The ANN architecture consists of four elements, namely, the number of layers, the number of neurons in each layer, the activation function of each layer and the training algorithm (Benardos & Vosniakos, 2007). The number of layer consists of three types of neuron layers, namely, the input, hidden and the output layers, while the number of neurons in each layer is specified by the number of input and output parameters that are used to model each problem. The activation functions are chosen based on the type of data such as binary, decimal, etc., and the type of layer. On the other hand, the training algorithm influences the training speed and performance of the ANN rather than the architecture itself (Benardos & Vosniakos, 2007).

2.4.2.4. Activation functions (AFs)

Activation functions (also known as transfer functions) defines the output of a neuron in terms of the linear combination of inputs (Mas & Flores, 2008). The activation function is used to limit the amplitude of the output of a neuron and is denoted as φ, which defines the output of a neuron in terms of the induced local field v (Haykin, 2009). There are different types of activation function including: rectified linear activation (ReLU), logistic (sigmond), threshold function, hyperbolic tangent (TanH) etc. Details of other types of AFs can be found

in Nwankpa et. al., (2018).

2.4.2.4.1. The sigmoid function (logistic function)

The sigmoid function (logistic function) is the most commonly used function (Mas & Flores, 2008), where the input to the function is transformed into a value between 0.0 and 1.0, with inputs larger than 1.0 transformed to the value 1.0 and values smaller than 0.0 snapped to 0.0 (Brownlee, 2019). The sigmoid function is defined as a strictly increasing function that exhibits a graceful balance between linear and nonlinear behavior with the shape of the function for all possible inputs as an S-shape (Haykin, 2009). The sigmoid function is defined as shown in equation (2.27) (Mas & Flores, 2008).

$$\varphi(v) = \frac{1}{1+e^{-av}} \qquad (2.\ 27)$$

where $a > 0$ is the slope parameter, the larger the value of a, the steeper the curve is.

2.4.2.4.2. The hyperbolic tangent function (TanH)

The hyperbolic tangent function (TanH) refers to the ratio between the hyperbolic sine and the cosine functions, with the range between -1 and +1 (Karlik & Olgac, 2010). The TanH function is preferred over the sigmoid function as models that used TanH are easier to train and often have better predictive performance (Brownlee, 2019). The TanH is denoted as shown in equation (2.28), where x is the input value, if x is negative, then a value 0.0 is returned, otherwise the value is returned.

$$\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \qquad (2.\ 28)$$

2.4.2.4.3. The rectified linear activation (ReLU)

The rectified linear activation (ReLU) function is the identity for positive arguments and zero otherwise (Clevert et. al., 2016) and is currently the most popular activation function for neural networks which was first proposed for restricted Boltzmann machines (Nair & Hinton, 2010) and later for neural networks (Glorot et. al., 2011). The ReLU is a simple calculation that returns the value provided as input directly or the value 0.0 if the input is 0.0 or less

(Brownlee, 2019) and is expressed as shown in equation (2.29).

$$f(x) = \max(0.0, x) \tag{2. 29}$$

2.4.2.5. Epoch

The number of epochs is a hyperparameter that defines the number of times that the learning algorithm will work through the entire training dataset and can be set to an integer value between one and infinity (Brownlee, 2018).

2.4.2.6. Types of ANN architectures

The basic ANN architectures are divided into feed-forward networks and recurrent networks. In feed-forward networks, the signal flow from the input to the output units, that is unidirectional in a feed-forward direction. The data processing may have a single layer (single layer neural network) or multiple layers (multi-layer neural network) with no feed-back connections. In the feed-forward networks, the transfer functions of the neurons do not affect the feed-forward behavior (Priddy & Keller, 2005). The advantage of feed-forward ANN is that they do not require a user specified problem solving algorithm but instead 'learn' from examples and possess an inherent generalization ability (Benardos & Vosniakos, 2007). The critical elements in a feed-forward neural network are the number of hidden layers and hidden neurons which affects the generalization ability of the neural network model (Benardos & Vosniakos, 2007). On the other hand, the recurrent networks contain feedback connections and may also be a single or multi-layered network. Recurrent networks are applied in nonlinear time series prediction, system identification and temporal pattern classification (Abraham, 2005).

2.4.2.7. ANN forecasting models

The advantages of ANN forecasting models are that they do not assume the underlining distribution for the data and they can extract the complex non-linear relationships in time series data. ANN forecasting models are categorized into two broad groups, namely:

1. Recursive multistep neural network (RMSNN) which is similar to ARIMA models in forecasting approach and has a single output node. A RMSNN model forecast one time step ahead and the network is applied recursively using previous predictions as inputs for subsequent forecasts (Mishra & Desai, 2006; Mishra et. al., 2007; Ornella, et. al., 2020).
2. Direct multistep neural network (DMSNN) is a type of network which has a single or multiple neurons in both the input and hidden layers and have several neurons in the output layer representing multiple month lead time forecasts (Mishra & Desai, 2006; Mishra et. al., 2007; Ornella et. al., 2020).

Both the RMSNN and DMSNN models use present index values and several months of past index values as inputs. Figure 2.2 shows the architectures of the RMSNN and the DMSNN (Ornella et. al., 2020).

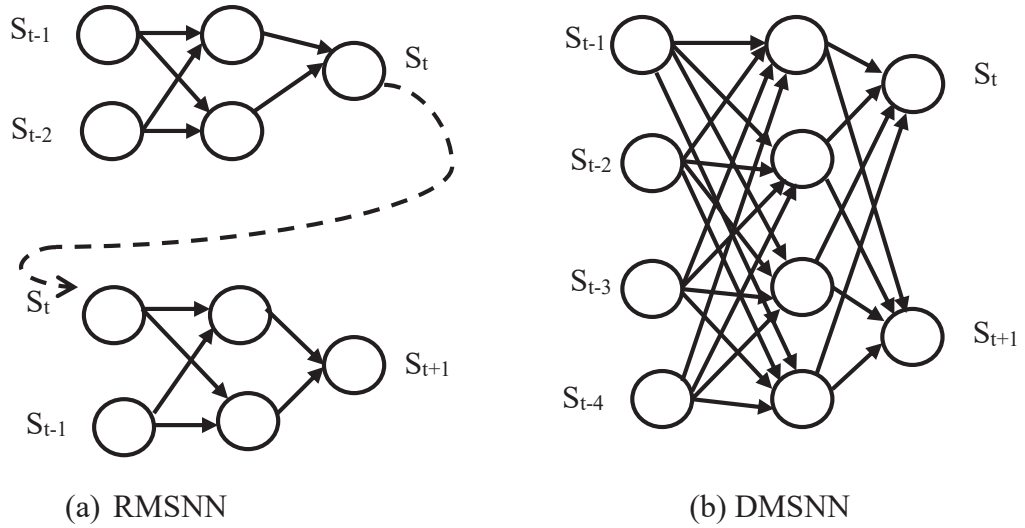

Figure 2. 2. Architectures of forecasting ANN models: (a) Recursive multistep neural network and (b) Direct multistep neural network

The multi-layer feed-forward neural network (MLP) using the direct multistep approach was used in this study which is explained in the section that follows.

2.4.2.8. Multilayer perceptron (MLP) model

A multilayer perceptron (MLP) is a feed-forward neural network and the most commonly used ANN which consists of a set of input units (input layer), one or more sets of computation nodes (the hidden layer) and one set of output nodes (the output layer) (Mas & Flores, 2008; Hamid et. al., 2011). Figure 2.3 shows the architectural graph of a MLP model with one hidden layer. MLP uses self-learning mechanisms by samples in the prediction of natural phenomena using environmental variables as the inputs of the network. The modeling starts with learning by samples and the transfer functions between the neurons are determined to achieve the most accurate weight of neurons (Jahani et. al., 2016). Each neuron receives the value of the weighted input variables periodically from other neurons and sends the new value to other processors, in addition, using a learning algorithm, the weights of neurons are adjusted continuously in the hidden layers to minimize the differences between target and network output values (Jahani et. al., 2020).

In a three-layer feedforward neural network, the input layer receives the data vector *x*, while the output layer gives the output vector, *y*. An activation function is applied to activate the neurons in the hidden layer, which can be expressed as shown in equation (2.30) (Ornella et. al., 2020).

$$y = \frac{1}{2}\left[\sum_{j=1}^{h}\left(w_j f_1 \sum_{i=0}^{n} w_{ji} x_i\right)\right] \qquad (2.30)$$

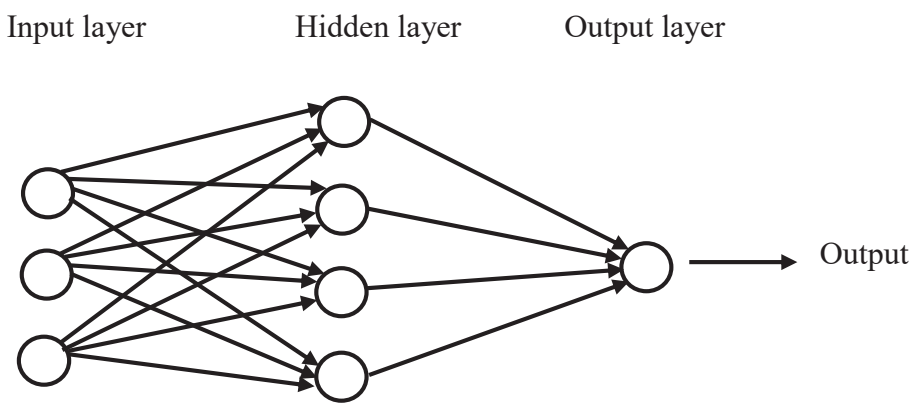

Figure 2. 3. Architectural graph of MLP with one hidden layer

2.4.3. The family of autoregressive moving average (ARMA) models

The family of ARMA models are flexible and among the most commonly used methods used to describe stationary time series. The models consist of the autoregressive (AR) part which describes how consecutive observations in time influence each other while the moving average (MA) parts capture possible unobserved shocks by allowing modeling of different phenomena observed in various fields (Guerrier et.al., 2019). The most commonly used family of ARMA models are explained in the sections that follows.

2.4.3.1. Autoregressive (AR) model

The AR model were first introduced by Yule (1926) and are a form of regression that links the observations of a particular moment with the values of previous observations at a specific time interval, with the term autoregressive indicating that it is a regression of a variable against itself (Ahmar & del Val, 2020; Hyndman & Athanasopoulos, 2018; NIST/SEMATECH, 2012). In an AR model, the forecast is done using a linear combination of past values of the variable. The AR model of order p can be written as in equation (2.31) (NIST/SEMATECH, 2012).

$$X_t = \delta + \phi_1 X_{t\text{-}1} + \cdots + \phi_p X_{t\text{-}p} + A_t \qquad (2.31)$$

Where X_t is the time series, μ is the process mean, A_t is the white noise and ϕ are the autoregressive parameters to be estimated, p is the order of the AR model and

$$\delta = \left(1 - \sum_{i=1}^{p} \phi_i\right)\mu \qquad (2.32)$$

2.4.3.2. Moving average (MA) model

The MA model was presented by Slutsky (1937), and uses the past forecast errors in a regression-like model, i.e., MA is a linear regression of the current value of the series against the white noise or random shocks of one or more prior values of the series (NIST/SEMATECH, 2012; Hyndman & Athanasopoulos, 2018). The MA model is expressed as shown in equation (2.33) (NIST/SEMATECH, 2012).

$$X_t = \mu + A_t - \theta_1 A_{t\text{-}1} - \cdots - \theta_q A_{t-q} \qquad (2.33)$$

where: X_t is the time series, μ is the mean of the series, $A_{t\text{-}i}$ are white noise terms and θ_1, ..., θ_q are the parameters of the model. The value of q is called the order of the MA model.

2.4.3.3. Autoregressive moving average (ARMA) model

Wold (1938) combined both the AR and MA models and showed that ARMA processes can be used to model all stationary time series as long as the appropriate order of *p*, the number of AR terms and *q,* the number of MA terms was appropriately specified (Makridakis & Hibon, 1998). The ARMA model was popularized by Box and Jenkins (1970) (thus, came to be known as the Box-Jenkins model) as tools for modelling and forecasting stochastic processes by developing a systematic methodology for identifying and estimating models that could incorporate AR and MA models (NIST/SEMATECH, 2012; Makridakis & Hibon, 1998). The ARMA model is expressed as in (2.34) (NIST/SEMATECH, 2012).

$$X_t = \delta + \phi_1 X_{t\text{-}1} + \phi_2 X_{t\text{-}2} + \ldots + \phi_p X_{t\text{-}p} + A_t - \theta_1 A_{t\text{-}1} - \theta_2 A_{t\text{-}2} - \cdots - \theta_q A_{t-q} \qquad (2.34)$$

Where the terms in the equation (2.34) have the same meaning as in the AR and MA models.

The Box-Jenkins model use the backshift operator *B,* which has the effect of changing time period *t*, to time period $t - 1$. Thus, $BX_t = X_{t-1}$ and $B^2 X_t = X_{t-2}$ (NCSS Statistical Software, 2021; Hyndman & Athanasopoulos, 2018). Hence, using the backshift operator, the ARMA model may be rewritten as shown in equation (2.35), which maybe abbreviated as shown in equation (2.36).

$$\left(1 - \phi_1 B - \cdots - \phi_p B^p \right) X_t = \left(1 - \theta_1 B - \cdots - \theta_q B^q \right) a_t \qquad (2.35)$$

$$\phi_p (B) X_t = \theta_q (B) a_t \qquad (2.36)$$

Where:

$$\phi_p\,(B) = \,(\,1 - \phi_1\,B - \cdots - \,\phi_p B^p) \tag{2. 37}$$

and

$$\theta_q\,(B) = \,(\,1 - \theta_1\,B - \cdots - \,\theta_q B^q) \tag{2. 38}$$

2.4.3.3.1. Stages in building a Box-Jenkins time series model

There are three development stages for a Box-Jenkins model, namely: identification, estimation and diagnostic checking. It is important to decide which family of models is to be fitted to the time series before starting the development stages to build the model. In the identification stage, the most suitable models to fit to the data can be selected by examining various types of graphs. Secondly, efficient estimates of the model parameters can be obtained during the estimation stage, by employing the method of maximum likelihood. Finally, the fitted models can be checked for any possible inadequacies during the diagnostic checking stage. In case the validation test reveals serious model anomalies for the fitted model which appears to be the most appropriate, then the necessary model modifications can be made by repeating the aforementioned three stages of model development (Hidel & McLeod, 1994). Figure 2.4 (Hidel & McLeod, 1994) shows the summary of the development stages of the Box-Jenkins model, and the stages are further explained in the sections that follows. Further details on the development stages can be found in Hidel & McLeod (1994).

2.4.3.3.1.1. Identification stage

The purpose of the identification stage is to determine if the series is stationary and if there is any significant seasonality that needs to be modeled. Stationarity can be assessed from a run sequence plot which should show constant location and scale and can also be detected using an autocorrelation plot. Non stationarity is often indicated by an autocorrelation plot with very slow decay. On the other hand, seasonality can be assessed from an autocorrelation plot, seasonal sub-series plot or a spectral plot. (NIST/SEMATECH, 2012; Hidel & McLeod, 1994). The autocorrelation plots are used to check the randomness in the time series data, while the seasonal sub-series plots detect the seasonality and is only useful if the period of

the seasonality is already known. On the other hand, the spectral plots are used to examine the cyclic structure in the frequency domain (NIST/SEMATECH, 2012; Hidel & McLeod, 1994).

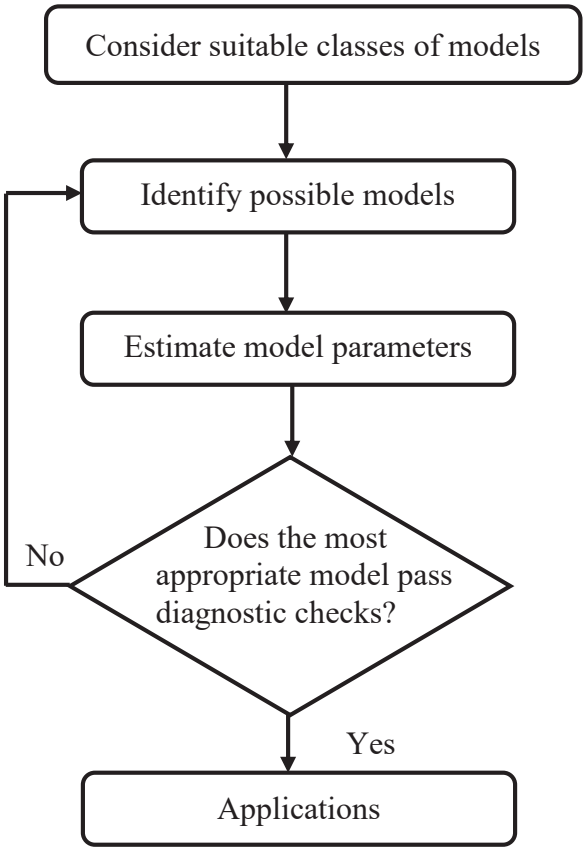

Figure 2. 4. Stages in the development of the Box-Jenkins model

In order to achieve stationarity, Box and Jenkins recommend the differencing approach. As aforementioned, the purpose of the identification stage is to detect seasonality and if it exists, to identify the order for the seasonal AR and seasonal MA parameters. It is advisable to apply a seasonal difference to the data and regenerate the autocorrelation and partial autocorrelation plots which may help in the model identification of the non-seasonal component of the model.

After addressing the stationarity and seasonality, the next step is to identify the order of the AR and MA parameters (p and q respectively) using autocorrelation and partial

autocorrelation plots. For an AR (1) process, the sample autocorrelation function (ACF) should have an exponentially decreasing appearance, although a mixture of exponentially decreasing and damped sinusoidal components may be observed for higher order AR process (NIST/SEMATECH, 2012). In the case of higher order AR processes, the sample autocorrelation plot should be supplemented with a partial autocorrelation plot, where, the partial autocorrelation of an AR (p) process becomes zero at lag $p + 1$ and greater. This is achieved by placing a 95% confidence interval on the sample partial autocorrelation plot and may be plotted using software or approximated using $\pm 2\sqrt{N}$, with N denoting the sample size (NIST/SEMATECH, 2012).

On the other hand, the autocorrelation function of a MA (q) process becomes zero at lag $q + 1$ and greater and can be examined using 95% confidence interval for the sample ACF on the sample autocorrelation plot (NIST/SEMATECH, 2012).

A mixed model approach is used when the ACF and partial autocorrelation function (PACF) does not work. For instance, in ARIMA *(p, d, q)* model, the ACF will be a mixture of exponential decay and damped sine waves after the first $q - p$ lags, while the PACF will have the same pattern after $p - q$ lags. Studying the first few correlations of each plots will assist to obtain the p and q terms. However, directly identifying the p and q terms using this method is difficult, thus a trial and error approach is used. To overcome this problem, recent methods such as final prediction error (FPE) and Akaike information criterion (AIC) are used to automate the model identification process and can be implemented using statistical software. Table 2.2 shows a summary of model identification using the sample ACF (NIST/SEMATECH, 2012).

2.4.3.3.1.2. Estimation stage

Estimation of the parameters for the Box-Jenkins model can be achieved using software that fits Box-Jenkins models. Various approaches such as non-linear least squares and maximum likelihood estimation (MLE) methods are used (NIST/SEMATECH, 2012). The maximum likelihood equation is solved using a nonlinear function maximization, while backcasting is

used to obtain estimates of the initial residuals. The estimation process is further explained in Box & Jenkins (1970).

2.4.3.3.1.1. Diagnostic checking

The final step after fitting the model is diagnostic checking, which is carried out by studying the autocorrelation plots of the residuals to check if large correlation values can be found. If the autocorrelation and partial autocorrelation are small, the model is considered adequate and forecasts are generated, however, if autocorrelations are large, the values of p and q are adjusted and the model is re-estimated (NCSS Statistical Software, 2021; Hidel & McLeod, 1994).

Table 2.2. Summary of model identification using the sample ACF.

Shape	Indicated model
Exponential, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to identify the order of the autoregressive model.
Alternating positive and negative, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to help identify the order.
One or more spikes, rest are essentially zero	Moving average model, order identified by where plot becomes zero.
Decay, starting after a few lags	Mixed autoregressive and moving average model.
All zero or close to zero	Data is essentially random.
High values at fixed intervals	Include seasonal autoregressive term.
No decay to zero	Series is not stationary.

2.4.3.4. Autoregressive integrate moving average (ARIMA) model

The ARMA model can be extended to non-stationary series by allowing the differencing of the data series, resulting in autoregressive integrate moving average (ARIMA) models, which allows each variable to be stated by its own lagged values and stochastic error terms (Durdu, 2010). In ARIMA model, the past observations are analyzed to formulate a model describing the inner correlation among them and the time series is extrapolated into the future according to the model (Han, et. al., 2010). ARIMA model is denoted as ARIMA (*p, d, q*) and can be written as shown in equation (2.39).

$$\varphi(B)\nabla^d X_t = \theta(B)a_t \qquad (2.39)$$

Where

$$\theta\,(B) = 1 - \theta_1\, B - \cdots - \theta_q B^q \qquad (2.40)$$

$$\varphi\,(B) = 1 - \varphi_1\, B - \cdots - \varphi_p B^p \qquad (2.41)$$

where, X_t is the observed series, θ_l and φ_l are model parameters, p and q are orders of the model. B is the backward shift operator ($B^k X_t = X_{t-k}$). Random errors a_t are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2 (white noise), ∇^d describes differencing operation to data series to make the data series stationary and d is the number of differencing.

2.4.3.5. Seasonal autoregressive moving average (SARIMA) model

The seasonal ARIMA (SARIMA) includes both the non-seasonal factors of the ARIMA model and seasonal factors in a multiplicative model, i.e., it uses differencing at a lag equal to the number of season (s) to remove additive seasonal effects (Hyndman & Athanasopoulos, 2018). The model is useful for modeling seasonal time series in which the mean and other statistics for a given season are not stationary across the years (Hipel & McLeod, 1994), i.e., when the time series data exhibits seasonality-periodic fluctuations that recur with approximately the same intensity each year. Thus, SARIMA model is suitable for forecasting

vegetation condition using monthly NDVI data since the NDVI values are subject to seasonal variations, with high NDVI values after a rainy season and low NDVI values in the dry season. In addition, the SARIMA model has relatively few parameters required to describe the time series and is popular due to its ease in implementation using statistical softwares, although, it is slower and requires more computational time compared to the exponential smoothing methods. Moreover, the model can only extract linear relationships in time series data and is not suitable for describing time series having stationarity of second order moments within each season across the years (Hipel & McLeod, 1994).

SARIMA model can grasp the historical information by:

1. Seasonal and regular differences in order to achieve stationarity.
2. Autoregressive (AR) which takes into account the past values.
3. Moving average (MA) which takes into account the current and previous residual series.

Hyperparameters are required to configure a SARIMA model for the trend and seasonal components of the time series. The trend component requires three parameters which are the same as the ARIMA model, that is: p: trend autoregression order, d: trend difference order and q: trend moving average order. The seasonal component requires four parameters that are not part of the ARIMA and includes the following elements: P: seasonal autoregressive order, D: seasonal difference order, Q: seasonal moving average order and s: the number of time steps for a seasonal period i.e., the number of observations per year (seasonality). The seasonality parameter influences the other parameters in the seasonal component (P, D and Q).

Thus, a SARIMA model is denoted as: SARIMA $(p, d, q) \times (P, D, Q)_s$, where: (p, d, q) is the non-seasonal part of the model and (P, D, Q) is the seasonal part of the model. The subscript s refers to the number of seasons per year. As mentioned above, the SARIMA $(p, d, q) \times (P, D, Q)_s$ model is referred to as a multiplicative model, because the non-seasonal and seasonal AR operators are multiplied together on the left hand side of the notation while the two MA operators are multiplied together on the right hand side. Moreover, the non-seasonal and seasonal differencing terms are multiplied together with the AR (Hipel &

McLeod, 1994).

If the data is stationary, non-seasonal or seasonal differencing is not required. A stationary SARIMA model is indicated as $(p, 0, q) \times (P, 0, Q)_s$ and such a model is sometimes known as seasonal autoregressive moving average (SARMA) model, which may be denoted as: $(p, q) \times (P, Q)_s$. For a SARIMA model which contains moving average (MA) parameters only, then the model is denoted as $(0, d, q) \times (0, D, Q)_s$, whereas if a model contains no MA parameters, then the SARIMA model is expressed as $(p, d, 0) \times (P, D, 0)_s$. In the case of a stationary non-seasonal model, the SARIMA model is written as ARMA (p, q) instead of SARMA $(p, q) \times (0, 0)_1$. Similarly, a non-stationary non-seasonal ARIMA model is denoted as ARIMA (p, d, q) rather than SARIMA $(p, d, q) \times (0, 0, 0)_1$ (Hipel & McLeod, 1994).

Using the back shift operator, the SARIMA model can be expressed as shown in equation (2.42).

$$\Theta_P(B^s)\theta_p(B)(1-B^s)^D(1-B)^d x_t = \Phi_Q(B^s)\phi_q(B)w_t \qquad (2.42)$$

where: Θ_P, θ_p, ϕ_q are polynomials of orders P, p, and q respectively.

2.4.3.5.1. Stages in constructing a SARIMA model

Similarly, to the Box-Jenkins model, the SARIMA model can be fitted to a given seasonal time series data using the identification, estimation and diagnostic checking as explained in section 2.4.3.3.1 and summarized in Figure 2.5. It is advisable to use at least seven years of seasonal data and at least 50 data points for a SARIMA model in order to get reasonable MLE for the model parameters (Hidel & McLeod, 1994). Table 2.3 shows a summary for the behavior of identification functions for SARIMA models (Hidel & McLeod, 1994). In addition to the ACF and PACF techniques, inverse autocorrelation function (IACF) and inverse partial autocorrelation function methods are used which are explained in details in Hipel & McLeod (1994).

Table 2.3. Summary for the behaviour of identification functions for SARIMA models

Function	Types of models		
	Pure AR	Pure MA	Mixed
	$(p, d, 0) \times (P, D, 0)_s$	$(0, d, q) \times (0, D, Q)_s$	$(p, d, q) \times (P, D, Q)_s$
ACF	Attenuates	Truncates after lag $q + sQ$	Attenuates
PACF	Truncates after lag $p + sP$	Attenuates	Attenuates
IACF	Truncates after lag $p + sP$	Attenuates	Attenuates
IPACF	Attenuates	Truncates after lag $q + sQ$	Attenuates

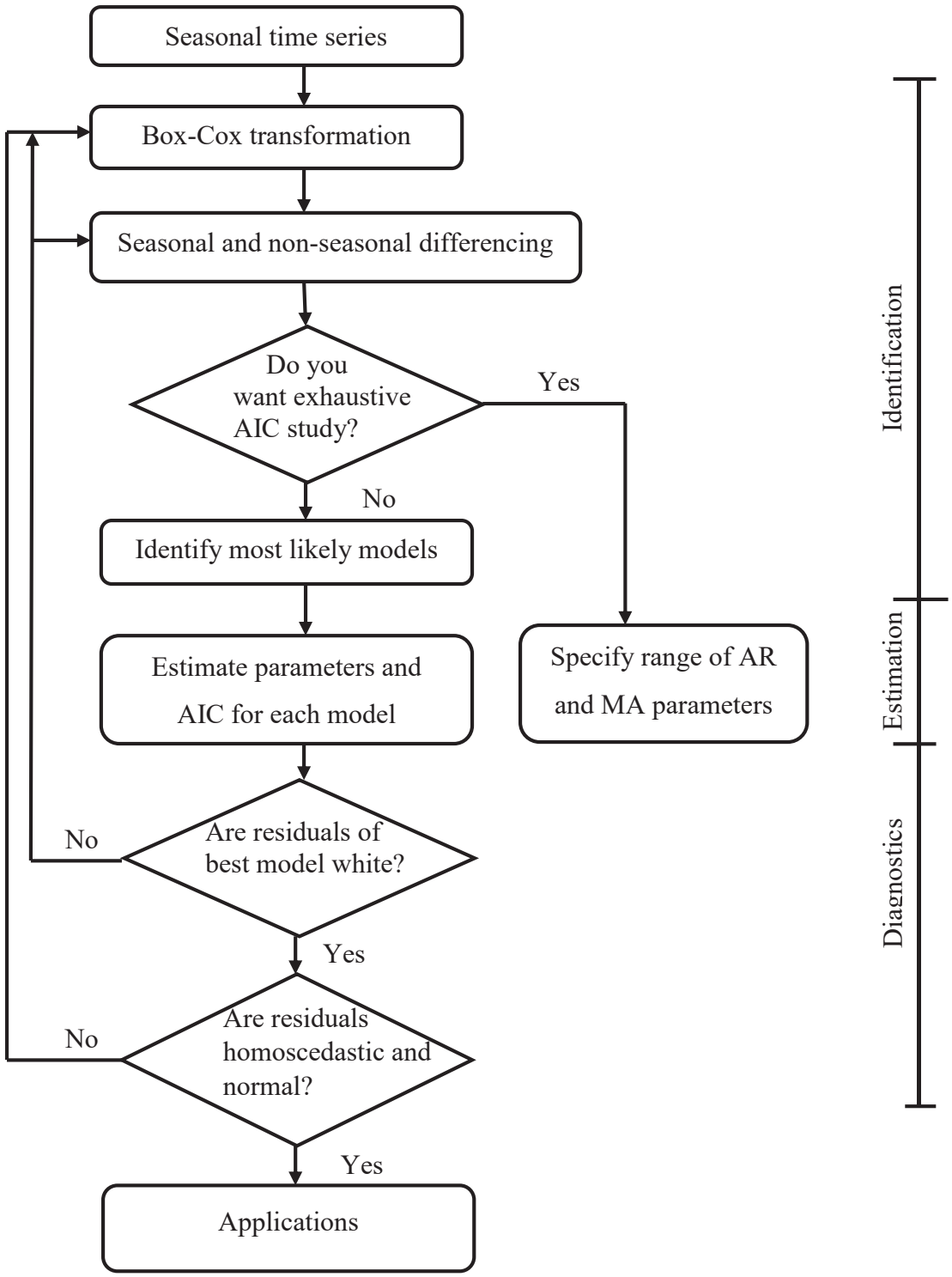

Figure 2. 5. Flowchart showing summary of the stages in the construction of a SARIMA model

CHAPTER THREE

MATERIALS AND METHODOLOGY

This chapter presents the study area, its location in east Africa and its characteristics. The following sections describe the data used, computational resources (virtual machine and supercomputer), method for determining models hyperparameters, implementation of the forecasting models and models validation metrics.

3.1 Study area

The area of interest in this research mainly focused on a section of the Middle Tana River Basin (MTRB) in Kenya (representing a low vegetation region) and sections of western and central Uganda and Ituri province in the Democratic Republic of Congo (DRC) (representing a high vegetation region) as shown in Figure 3.1 and Figure 3.2. Figure 3.1 shows the location of the study area while Figure 3.2 shows the satellite map showing the difference in vegetation cover between the two regions. Both regions have an area of 22, 500 km^2 equivalent to 600 $\times$ 600 pixels. The regions receive bimodal rainfall seasons, occurring between March to May and October to December (Townshend & Justice, 1986; Chapman & Chapman, 2004). Table 3.1 shows the characteristics of the study area.

The low vegetation region is part of the Tana River Basin in Kenya which covers 22% of the country's total land mass (an area of 126, 000 km^2) and is home to 18% of the country's population (Beukering et. al., 2015; Agwata, 2005). The Basin has a rich and diverse ecosystem including montane, grasslands, mangroves, wetlands, coastal forests and cropland (Beukering et. al., 2015; Wetlands International Kenya, 2015). The MTRB is characterized as an arid and semiarid land (ASAL) and lies at coordinates 0º to 2º south and 38º to 40º east (Figure 3.1 (a)). Sections of Kitui, Garissa and Tana River counties in Kenya, form the MTRB region used in this study. The region receives an annual mean rainfall below 800 mm and annual average temperature between 30 ºC to 33 ºC (Baker et. al., 2015; Kipchirchir, 2014). The region is extensively used for livestock, agriculture, tree harvesting

and uncontrolled sand mining leading to removal of trees without regeneration (Knoop et. al., 2012).

The high vegetation region has a tropical climate and lies at coordinates 0º to 2º north and 30º to 32º east (Figure 3.1 (a)). The region receives an annual average rainfall of 1500 mm and an average temperature ranging between 16 ºC to 25 ºC (Byrnes, 1992). The main economic activity in Western and Central Uganda is agriculture, which is practiced for food availability and a means for cash generation (Wichern et. al., 2017). Similarly, the main economic activity in Ituri is agriculture (both for staple and for cash crop production), as well as livestock production, mining and charcoal (Douma, et. al., 2009).

Table 3. 1. Characteristics of the study area

Region	Elevation	Type of vegetation
Kitui	400 m to 1800 m a. s. l	Scrublands, wooded bushlands
Garissa	20 m to 400 m a. s. l	Shrub savannah and woody trees
Tana River	0 m to 200 m a. s. l	Scrubland, thorny thickets, shrubs, grasses
Central Uganda	1200 m to 1500 m a. s. l	Wooded savannah
Western Uganda	1200 m to 1500 m a. s. l	Long grass and forest
Ituri province	1200 m to 1500 m a. s. l	Tropical forest, savannah, giant evergreen trees, thick woody vines, short trees, bushes and ferns

(Source: Verdcourt et. al., 1975; Malonza et. al., 2006; Bélair et. al., 2010; Bernard, 1998; County Government of Tana River, 2018; Obua et. al., 2010)

Figure 3.1 shows the map of the study area. The lower right map shows the low vegetation region located in the Middle Tana River Basin (comprising sections of Garissa, Kitui and Tana River counties) in Kenya. The lower left hand map shows the high vegetation region located in sections of Ituri Province in Democratic Republic of Congo, western and central Uganda. The inset map on top shows the east Africa map and the MODIS tiles used

in this research. The MODIS h21v08 tile was used to extract pixels for the high vegetation region while the h21v09 tile was used to extract pixels for the low vegetation region.

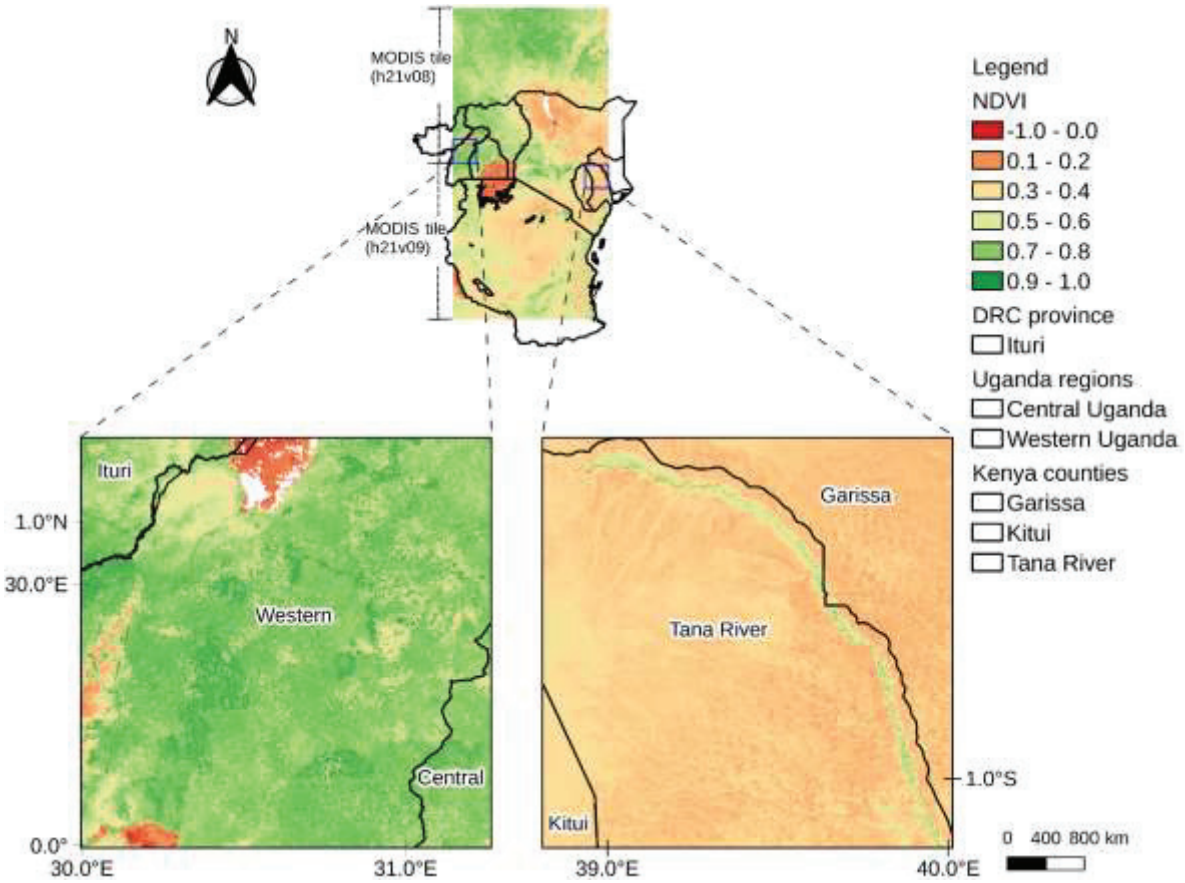

Figure 3.1. Map of the low vegetation region (Middle Tana River Basin) and high vegetation region (Western and Central Uganda and Ituri Province in DRC) (Source: Author).

Figure 3.2 shows the satellite map of the study area. The inset map on the top shows the location of the study area in Kenya (Garissa, Kitui and Tana River counties), Uganda (western and central regions) Democratic Republic of Congo (Ituri Province). The lower right map shows the low vegetation region, while the lower left map shows the high vegetation region.

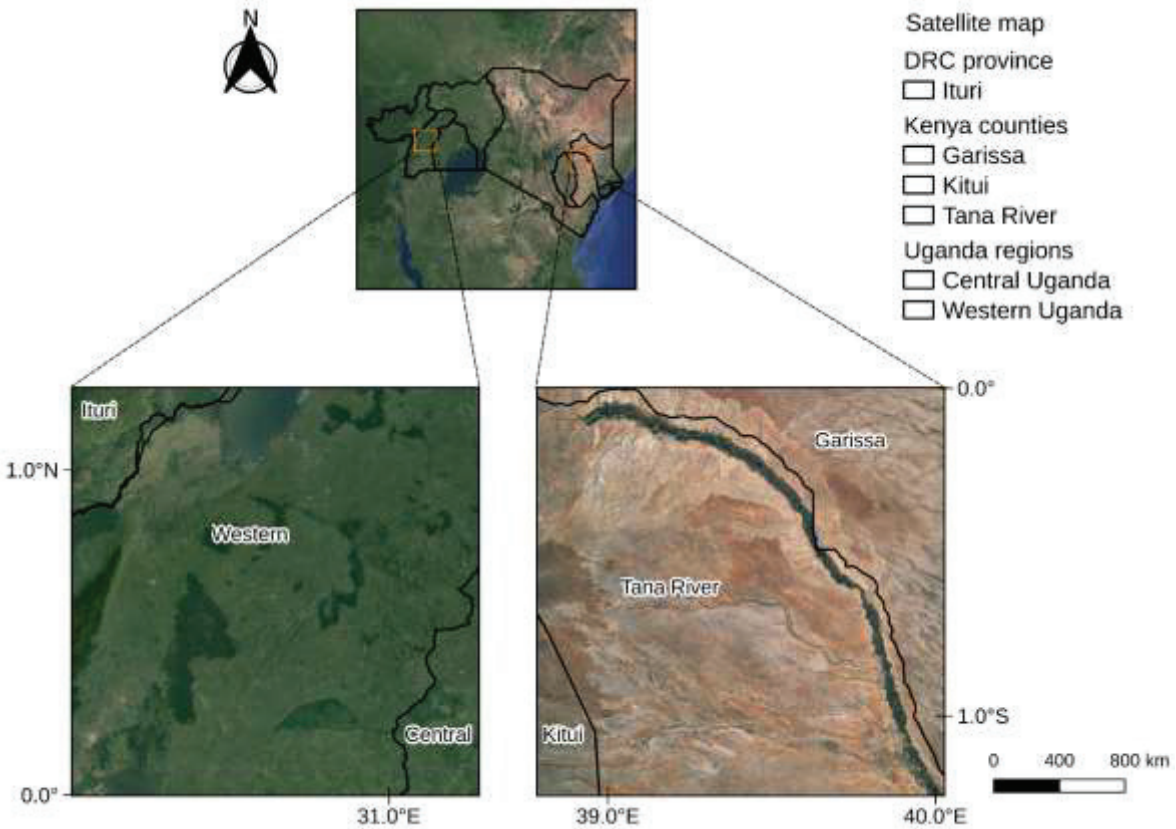

Figure 3.2. Satellite map of the low vegetation region (Middle Tana River Basin) and high vegetation region (Western and Central Uganda and Ituri Province in DRC) (Source: Author; Satellite map source: Google).

3.2 The normalized difference vegetation index data

The MODIS/Terra VI (MOD13Q1 v006 product) with 250 m spatial resolution and 16 – days composite was obtained from the Land Processes Distributed Active Archive Center (LP DAAC) using the Earthdata Search tool (Earthdata Search, 2019). The MOD13Q1 product contains 12 scientific datasets (SDSs) namely: 250 m 16 days NDVI, 250 m 16 days EVI, 250 m 16 days VI Quality, 250 m 16 days red reflectance, 250 m 16 days NIR reflectance, 250 m 16 days blue reflectance, 250 m 16 days MIR reflectance, 250 m 16 days view zenith angle, 250 m 16 days sun zenith angle, 250 m 16 days relative azimuth angle, 250 m 16 days composite days of the year and 250 m 16 days pixel reliability (Didan, 2015).

In this research, the first layer (250 m 16 days NDVI) was extracted to monitor

vegetation condition in the study area. For each region, 452 images covering a period of approximately 19 years, from February 2000 to September 2019 was used. The images were extracted from sections in the h21v09 and h21v08 tiles to represent the low and high vegetation regions respectively (Figure 3.1). As mentioned before, the top right corner of the h21v09 tile was used to extract pixel-wise NDVI data for the low vegetation region while the bottom left corner of the h21v08 tile was used to extract pixel-wise NDVI data for the high vegetation region. To cover the study area in the low and high vegetation regions, 600 × 600 pixels was extracted in both tiles (h21v08 and h21v09).

Due to the large number of pixels (600 × 600 = 360, 000 pixels) covering the study area, it was challenging to perform the temporal analysis of NDVI using all the pixels in the two regions as plotting the 600 × 600 pixels would have resulted in 360, 000 graphs (1 graph for each pixel), which would have made it cumbersome to perform NDVI time series analysis as well as to study the temporal pattern. Moreover, using the entire 600 × 600 pixels to perform a grid search for the models' hyperparameters and to develop the models would have required higher computational power, costly and time consuming. Thus, to address these challenges, 1 × 1 pixel (subset data) was extracted from each tile in both regions and was used to analyze the temporal pattern of the NDVI time series data. The first pixel in the top right corner of the h21v09 tile and the first pixel from the bottom left corner of the h21v08 tile were extracted to perform the time series analysis. In addition, the subset data was also used to grid search the models' hyperparameters and develop the models, since pixel-wise modeling requires high computational power and are costly. Hence, it is recommended to use a subset of the data to develop the models and then use the developed models to the entire dataset. and used to train and test the models developed using the subset data. After developing the models using the subset data, the models were trained and tested using the entire data (600 × 600 pixels) to cover the entire study area.

The retrieved images were in the hierarchical data format (HDF), sinusoidal tile grid projection and the file names contained the Julian date of acquisition, which was useful when creating the NDVI time series. Jupyter notebook was used in order to integrate coding,

monitor the computational output and visual the plots from the NDVI time series data. The notebook was useful to extract the 1 × 1 pixel, perform temporal analysis, grid search models' hyperparameters and to develop the models. However, extracting the 600 × 600 pixels, training and testing the models for the entire study area was performed on the Linux command line (also known as terminal/shell/console) on the Oracle's VM and supercomputer. This was because, the saving and loading of the data back into the memory using Jupyter notebook would have been slower and time consuming when performed using a laptop.

The first step was to load the libraries needed to perform the NDVI time series analysis. The GDAL library was used to enable working with the HDF files, while the numpy package was used to work with arrays. Pandas (McKinney, 2010) library was used to create the dataframe and save the data into CSV file. The calendar module was used to convert the Julian dates to Gregorian dates. Due to high computational power of pixel-wise modeling, parallel computing was implemented, which was achieved using concurrent.futures module.

After loading the libraries, the date format was converted from Julian to Gregorian calendar as the Gregorian calendar is the most widely used and accurate calendar. The next step was to extract the pixel(s) using GDAL library in order to open the HDF files, examine and list the subdatasets (SDSs) in the MODIS MOD13Q1 data, and then load the NDVI subdataset. The pixel-wise NDVI data was then converted into a 2D numeric array. The final step was to save the pixel-wise NDVI dataframe into CSV file since the HDF files have slower loading speed compared to CSV files. In addition, CSV doesn't require additional memory to save or load data. The CSV file contained 452 rows representing all the NDVI images extracted during the study period (February 2000 – September 2019). The 1 × 1 pixel CSV file contained 2 columns, with the first column representing the date and the second column representing the NDVI value. Whereas, the 600 × 600 pixels CSV file contained 360, 001 columns, with the first column representing the date and the rest of the columns representing the NDVI values for each pixel. The valid range for the NDVI extracted from the MOD13Q1 product was -2000 to 10000, however, traditionally NDVI values range between -1 to +1, thus, a scale factor of 0.0001 (Didan, 2015) was applied to rescale the extracted pixel-wise NDVI data between -1 and +1. The mean of all 16 – days composite

pixel-wise NDVI in every month was calculated to provide monthly mean pixel-wise NDVI data. After resampling using the mean, the final CSV file contained 236 rows, 2 columns for 1 × 1 pixels and 360, 000 columns for the 600 × 600 pixels.

All scripts used in data extraction, grid search and model development as well as the time series analysis were written in Python programming language (Van Rossum & Drake, 2009) and the OS used was Ubuntu 16. 04.

3.3 Virtual machine

Pixel-wise modelling requires huge computational efforts (Mutti et. al., 2020) thus, a virtual machine (VM) from Oracle Cloud Infrastructure (OCI) was used to perform this research. The VM provides flexible image management and supports various operating systems (OS) such as Microsoft Windows Servers and enterprise Linux OS such as CentOS, Oracle Linux and Ubuntu. The VM enables deployment of applications in multi-region, multi-Availability Domain (AD) and multi-fault domain configurations and provides a secure and flexible network at a lower cost. The VM also caters for a low latency block storage service which provides a highly available, persistent, network-attached storage volume optimized to deliver low latency and high Input/Output Operations Per Second (IOPS). Oracle block volume provides versatile system boot disks that can be backed up and restored to new volumes (Oracle, 2020).

The VM instance provides different shapes. This study used the VM.Standard2.8, which has 8 OCPUs (Oracle Compute Unit) and 120 gigabytes (GB) of memory. The instance shape supports up to 1 Petabyte (PB) of remote block volume. Ubuntu was used as the OS and the platform image was Canonical-Ubuntu-16.04-2020.02.18-0. Table 3.2 shows the summary of the specifications of the VM.

Table 3. 2. Summary of the specifications of of the virtual machine from Oracle Cloud Infrastructure (OCI)

Compute shape	VM.Standard2.8
Oracle Compute Unit (OCPU)	8
Processor	Intel Xeon Platinum 8167M
Base frequency	2.0 GHz
Maximum turbo frequency	2.4 GHz
Maximum virtual network interface cards (VNICs) Total: Linux	8
Memory	120 gigabytes (GB)
Remote block volume	1 Petabyte (PB)
Operating systems (OS)	Ubuntu
Platform image	Canonical-Ubuntu-16.04-2020.02.18-0
Maximum Network Bandwidth	8.2 Gbps

3.4 Hyperparameters optimization

Hyperparameters are higher level properties of the algorithm statistical model which strongly influence its complexity, speed in learning, application results (Chicco, 2017) and are important to minimize the loss function. Hyperparameters are key to produce accurate model classification or prediction (Dogo et. al., 2018; Roy, 2021; Jamieson & Talwalkar, 2016; Brownlee, 2020). A model's hyperparameter is external to the model and cannot be estimated from the data, thus, various approaches are used such as grid search, manual search, random search etc.

Grid search is the most widely used approach for hyperparameter optimization due to its simplicity in implementation and being the default choice for open source machine learning packages (Bergstra & Bengio, 2012; Chicco, 2017; Jamieson & Talwalkar, 2016).

The approach methodically builds and evaluate a model for each combination of algorithm parameter specified in a grid (Liashchynskyi & Liashchynskyi, 2019).

The grid search approach was used in this research and was implemented using the GridSearchCV class from scikit-learn library which performs an exhaustive search and generates candidates from a specified grid parameter values (Pedregosa, et. al., 2011). Tables 3.3, 3.4 and 3.5 shows the search space used in the grid search for the hyperparameters for Holt-Winters, MLP and SARIMA models used in the research.

Table 3. 3. Search space for hyperparameters for the Holt-Winters model

Trend component	Additive (add), multiplicative (mul)
Seasonal component	Add, mul
Seasonal periods	12 (for 12 months yearly data)

Table 3. 4. Search space for hyperparameters for the MLP model

Optimizer search space	Adam, Adamax, Nadam
Activation search space	ReLU, TanH, linear
Number of epochs	2000, 3000, 4000, 5000

Table 3. 5. Search space for hyperparameters for the SARIMA model

Trend order	0, 1, 2
Seasonal order	0, 1, 2
Trend	No trend, constant, linear and constant with linear trend
Seasonal length	12 (for 12 months yearly data)

3.5 Implementation of the forecasting models

The section below describes the experimental procedure used to implement the models used in this research.

3.5.1 Implementation of the Holt-Winters model

This section is based on: Omar M. S. and Kawamukai H. (2021). Prediction of NDVI using the Holt-Winters model in high and low vegetation regions: A case study of east Africa. Scientific African, 14, e01020, pp. 1 - 10.

Since the pixel-wise NDVI time series data exhibited both trend and seasonal variation, the Holt-Winters model was used in this research instead of the SES and double exponential smoothing models. The Holt-Winters model was implemented using the 'statsmodels.tsa.holtwinters.ExponentialSmoothing' class from the statsmodels library; a python module for statistical analysis and statistical data exploration (Skipper & Perktold, 2010). The model was automatically tuned to calculate the optimal values for the α, β and γ. The results from the grid search returned the hyperparameters for the trend, seasonal and seasonal periods (number of time steps) as add, add and 12 (for 12 months' yearly data) respectively, resulting in an additive model. The dataset was split (based on temporal sampling) into training and testing sets at a ratio of 80% to 20% respectively. The training set was used to develop the model while the testing set was used to evaluate the model's performance. The 'fit()' function was used to fit the model on the training set and the 'forecast()' function was used to make the predictions. Detailed explanation of the functions and classes can be found in the statsmodels module (Skipper & Perktold, 2010). The predicted pixel-wise NDVI data in CSV format was then used to create spatial maps for the low and high vegetation regions using matplotlib library (Hunter, 2007).

The model was first trained on a sample size of 10 × 10 pixels and 100 × 100 pixels for the low and high vegetation regions. The model was then retrained using the same hyperparameters for the 600 × 600 pixels (total area for the study region) to get the 6 months

forecast pixel-wise NDVI data. Forecast evaluation metrics such as MAPE, RMSE and MAE were computed to evaluate how close the predicted values are to the true values.

3.5.2 Implementation of the SARIMA model

This section is based on: Omar M. S. and Kawamukai H. (2021). Comparison between the Holt-Winters and SARIMA models in the prediction of NDVI in an arid region in Kenya using pixel-wise NDVI time series. Academic Journal of Research and Scientific Publishing, 2 (23), pp. 1 – 15.

Similarly, to the Holt-Winters model, the SARIMA model was implemented using the 'SARIMAX' class from the statsmodel library. The trend and seasonal hyperparameters were defined using the 'order' and 'seasonal_order' arguments respectively and their configurations was obtained from the results of the grid search. The order (1, 0, 1) and seasonal_order = (0, 0, 2, 12) was used to define the model. The order argument defines the AR parameters, differences and MA parameters of the model as 1, 0, and 1 respectively, while the seasonal_order argument denotes the seasonal component ($(P, D, Q)_s$) of the model, with the length of the seasonal period (s) as 12, for the 12 months' yearly data.

Similarly, to the Holt-Winters model, the dataset was split (based on temporal sampling) into training and testing sets at a ratio of 80% to 20% respectively. The training set was used to develop the SARIMA model while the testing set was used to evaluate the model's performance. The 'fit()' function was used to fit the model on the training set and the 'forecast()' function was used to make the predictions. Detailed explanation of the functions and classes can be found in the statsmodels module (Skipper & Perktold, 2010).

The SARIMA model was first trained on a sample size of 1 × 1 pixel in the low vegetation region and later retrained for 600 × 600 pixels for the low vegetation region. Forecast evaluation metrics was computed to evaluate the performance of the model. Due to time, budget constraints and limitations of the computational resources, 100 × 100 pixels was used for the high vegetation region.

3.5.3 Implementation of the MLP models

This research employed the feed-forward multi-layer network using the direct multistep approach. The MLP models were developed by comparing the loss function of different MLP models from the grid search using various activation functions, number of epochs and optimization algorithms. Three MLP models which had the best score from the grid search were developed: MLP 1 (3000 epochs, ReLU, Adam), MLP 2 (4000 epochs, TanH, Adamax) and MLP 3 (5000 epochs, ReLU, Nadam). As mentioned in section 2.4.2.4., the activation function determines the relationship between the input and output nodes. In this research, two activation functions consisting of ReLU and TanH were used.

MLP models are slow and expensive to train (Fung et. al., 2020; Hansson et. al., 2016), thus, to reduce the computational burden in the model development, a small subset of the dataset (using 1 × 1 pixels) was fitted to speed up the search for the models' hyperparameters as explained in section 3.2. After the hyperparameters were determined, the entire 600 × 600 pixels for the entire study area was fed to the network for 6 months lead pixel-wise NDVI prediction. The MLP models were implemented in Keras (Chollet et. al., 2017), a deep learning library for python, running on top of the machine learning platform: TensorFlow (Abadi et. al., 2015).

Since the MLP models are slow to train and are thus, time consuming, a supercomputer from the Research Institute for Information Technology, Kyushu University was used instead of Oracle's VM. The supercomputer system is based on Intel Xeon Gold (Skylake architecture) CPUs and NDIVIA Tesla P100 (Pascal architecture) GPUs, with peak performance around 10 PFLOPS. Table 3.6 shows an overview of the specifications for the subsystem B used in this research. More details on the supercomputer can be found at: https://www.cc.kyushu-u.ac.jp/scp/eng/system/ITO/

Table 3. 6. Summary of the specifications of Subsystem B of the supercomputer from the Research Institute for Information Technology, Kyushu University

Machine		Fujitsu PRIMERGY CX2570 M4
	Central Processing Unit (CPU)	Intel Xeon Gold 6140 (Skylake-SP) (2.3 GHz (Turbo 3.7 GHz), 18 core） x 2 / node
	Graphics Processing Unit (GPU)	NVIDIA Tesla P100 (Pascal) (1,328 - 1,480 MHz, 56 SM (3584 CUDA core)) x 4 / node
Computing node	Theoretical Peak Performance	CPU : 2,649.6 GFLOPS / node (Double Precision) GPU : 5.3 TFLOPS (with Boost Clock) / 1GPU (Double Precision)
	Amount of Memory	Double Data Rate 4 (DDR4): 384 GB / node High bandwidth Memory 2 (HBM2): 16 GB / 1 GPU
	Memory Bandwidth	DDR4 : 255.9 GB/sec / node HBM2 : 732 GB/sec / 1GPU
	CPU-GPU Connection	Peripheral Component Interconnect Express (PCI-Express) Gen.3 x16 (16GB/sec)
	GPU-GPU Connection	NVLink (20GB/sec x 1or2)
Number of Nodes		128
Total Number of Cores		CPU : 4,608 GPU : 1,835,008

Total Theoretical Peak Performance	CPU : 0.34 PFLOPS (Double Precision) GPU : 2.71 PFLOPS (Double Precision)
Total Amount of Memory	DDR4 : 49 TB HBM2 : 8.19 TB
Interconnect	InfiniBand Enhanced Data Rate (EDR) 4x (100Gbps)
Local Storage of Node	1TB Hard disk drive (HDD) x 2 Serial advanced technology attachment (SATA) 800GB HDD x 1 (SATA)

The network used in this study consisted of 24 nodes in the input layer, 100 nodes in the hidden layer and 6 nodes in the output layer (indicating a one to six-month lead time), represented using the notation: 24/100/6. The input nodes corresponded to the number of lagged observations of the monthly mean pixel-wise NDVI data which were used to discover underlying pattern in the pixel-wise NDVI time series to predict future NDVI values. The nodes in the hidden layer allowed the network to detect the feature, capture the pattern in the data and perform nonlinear mapping between the input and output variables (Mishra & Desai, 2006; Mishra et. al., 2007). This research used one hidden layer as it has been proved to be sufficient for ANN to approximate any complex nonlinear function with desired accuracy (Hornik et. al., 1989). 24 previous monthly mean pixel-wise NDVI values were tested for optimal NDVI series within this range.

The network was trained for 3000, 4000 and 5000 epochs using Adam, Adamax and Nadam optimizers for MLP 1, MLP 2 and MLP 3 respectively and optimized using the MSE loss function using back propagation algorithm. The default values for the optimizers provided by Keras (Chollet et. al., 2017) were used. Further details on defaults values of Adam, Adamax and Nadam optimizers can be found on Keras (Chollet et. al., 2017).

After fitting the model to the training set, the model was used to make predictions by giving the last sequence of the input data, which was reshaped into a 2-D array before making

the predictions using the 'predict()' function. The above process was repeated for each pixel to get the pixel-wise NDVI prediction for 6-month lead for the study area. The performance of the predictions resulting from the MLP models was evaluated using MSE, RMSE and MAE. The predicted pixel-wise NDVI data in CSV format was then used to create spatial maps using matplotlib library (Hunter, 2007). The NDVI values were classified in the range between –1 to +1, with -1 representing a water body and +1 indicating high NDVI value.

3.6 Models validation metrics

The testing dataset was used as the input to evaluate the performance of the Holt-Winters, SARIMA and MLP models using statistical indicators, namely: MAE, MAPE, MSE and RMSE.

The metrics were implemented using scikit-learn, a python module that provides tools for predictive data analysis (Pedregosa, et.al. 2011). All metrics were imported from the 'sklearn.metrics' module, which implements functions for assessing prediction errors (Pedregosa, et.al. 2011). The following functions from the 'sklearn.metrics' module were used to implement the above mentioned metrics:

1. 'mean_squared_error' function computed MSE and RMSE.
2. 'mean_absolute_error' function computed MAE.
3. 'mean_absolute_percentage_error' function computed MAPE.

The evaluation metrics are described in the sections that follows. The symbols n, $\hat{y}_i$ and y_i used in the equations (3.1) to (3.4) refers to: n is the number of observations, y_i is the observed value and $\hat{y}_i$ is the predicted value.

3.6.1. Root mean square error

Root mean square error (RMSE) shows the difference between the forecasted and observed values. It represents the sample standard deviation of the difference between predicted and observed values and measures the overall performance across the entire dataset and provides a good measure of the model. RMSE varies from zero for perfect estimates to large positive

values for poor estimates (Neupane et. al., 2021). RMSE is computed as shown in equation (3.1).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (3.1)$$

3.6.2. Mean absolute error

Mean absolute error (MAE) is the mean absolute value of the absolute errors and is used to measure how close the forecast values are to the observed values (Neupane et. al., 2021). MAE is a risk metric corresponding to the expected value of the absolute error loss (Pedregosa, et.al. 2011). The MAE value zero represents a perfect fit with the values always greater than or equal to zero (Sharma & Deo, 2021). MAE is computed as indicated in equation (3.2).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (3.2)$$

3.6.3. Mean square error

Mean square error (MSE) uses the values of the predictor variables associated with the future observations and the magnitude of the estimated variance (Allen, 1971). MSE is a risk metric corresponding to the expected value of the squared (quadratic) error (Pedregosa, et.al. 2011). Equation (3.3) was used to calculate MSE.

$$MSE = \frac{1}{n}\sum_{i}^{n}(y_i - \hat{y}_i)^2 \qquad (3.3)$$

3.6.4. Mean absolute percentage error

Mean absolute percentage error (MAPE), is also known as mean absolute percentage deviation (MAPD) and refers to the average of the absolute percentage error of forecasted values with smaller values indicating better forecast results (Sharma & Deo, 2021). MAPE is calculated as shown in equation (3.4).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \qquad (3.4)$$

CHAPTER FOUR

RESULTS AND DISCUSSIONS

Chapter four presents the research findings on the forecasting of vegetation condition in east Africa (low and high vegetation regions) and prediction of NDVI using pixel-wise NDVI data, stochastic and ANN models. The following sections provides the temporal, spatial, statistical and comparative analysis of the models for the low and high vegetation regions. These findings were accepted for publication in peer reviewed journals titled Scientific African Journal and Academic Journal of Research and Scientific Publishing as mentioned in Chapter three and part of the results were presented during the 6th International Conference on Energy and Environmental Science (ICEES 2022) and Annual Research Presentation Conference on Spatial Information Science (CSIS DAYS 2021).

4.1 Introduction

Vegetation plays an important role in protecting the land surface from erosion, preventing desertification and conserving soil and water (Sun et. al., 2015). Moreover, vegetation covers such as forests supply water, provide livelihoods, mitigate climate change and are essential for sustainable food production (FAO & UNEP, 2020). However, vegetation cover is under threat due to human activities such as extensive use of land for agriculture, overgrazing and deforestation which are driven by a rapid population growth. Nearly 5 million hectares of natural vegetation is lost per year. The east Africa region has seen an increase in agricultural land to 57% between 1975 – 2000 at the expense of natural vegetation, which decreased by 21% (Brink & Eva, 2009). In addition, there is insufficient monitoring of vegetation with 21% of the land experiencing a decline in vegetation cover (Waswa, 2012) and a decrease in natural vegetation formations such as forests, bushlands, grasslands and wetlands (Githui et. al., 2009). Thus, monitoring the temporal changes and predicting future trends of NDVI is very important for understanding changes in vegetation condition which may assists in resource allocation and planning, especially in the low vegetation region, which is characterized by extensive use of land for livestock grazing.

Pixel-wise NDVI time series was used to develop forecasting models using MODIS NDVI data extracted from MOD13Q1 product. Pixel-wise NDVI data was used for its robustness, although it required higher computational power. The monthly mean NDVI time series was derived from the average values of NDVI data in each month during the study period. Although NDVI is affected by other factors such as rainfall, temperature, soil moisture etc, this research demonstrated that it is possible to monitor and forecast vegetation condition in a region as NDVI can be forecasted by taking into consideration its past values alone (Mutti, et. al., 2020).

In this study, the Holt-Winters and SARIMA models with 12 months' seasonal period was applied as NDVI is influenced by factors that have an annual seasonal pattern such as seasonal precipitation. All models (Holt-Winters, MLP and SARIMA) were developed using the optimal values obtained from the grid search to ensure that the models with the best accuracy were used.

The models' performance was evaluated using the RMSE, MSE, MAPE and MAE, which are widely used statistical metrics in vegetation studies to evaluate discrepancies between the actual and predicted data. Lower values of the aforementioned statistical metrics correspond to a model with better predictive performance. Moreover, a small difference between MAE and RMSE means a smaller variance in the individual errors in the data. Forecasted pixel-wise NDVI spatial maps at 6 months lead were created from the forecasting results using matplotlib library (Hunter, 2007).

4.2 Temporal analysis using 1 × 1 pixel NDVI data in the low and high vegetation regions

Time series analysis in vegetation monitoring provides an opportunity to predict future NDVI values based on lagged NDVI values. The temporal trends of 1 × 1 pixel's monthly mean NDVI from February 2000 to September 2019 for the low and high vegetation regions are shown in Figure 4.1 and 4.2 respectively. As mentioned in section 3.2, the first pixel from

the top right corner of h21v09 and the first pixel from the bottom left corner of h21v08 MODIS tiles were selected for the low and high vegetation regions respectively.

The monthly mean NDVI showed a temporal pattern with distinct low and peak values in both regions. High NDVI values were observed in the high vegetation region compared to the low vegetation region which may be attributed to the slower greening speed in arid regions (Yao et. al., 2019), and is mainly driven by precipitation anomalies (Adler et. al., 2017; Fensholt & Rasmussen 2011; Gu et. al., 2015). In both regions, NDVI was marked by low and peak values with monthly variations throughout the study period. This trend can be attributed to seasonal climate patterns, with low NDVI values during the dry season and high NDVI values during the rainy season. Similar observation was found in previous studies such as Kinyanjui (2010). In both regions, the year 2011 was marked by low NDVI values which may be attributed to the worst drought that occurred in east Africa due to the the failures of both the preceding rainy seasons and the prevailing La Niña conditions in late 2010 and early 2011 (AghaKouchak 2015; FEWS NET 2011; Lott et. al., 2013). Moreover, the failure in the rainy seasons led to crop failures (Funk 2011). Notably, the low vegetation region recorded an NDVI value of approximately 0.2 as plants were sparser or growing less than average pointing to poor plant growth during the growing season (NASA Earth Observatory, 2011). On the other hand, the high vegetation region had an NDVI around 0.4 indicating good plant growth due to leafier photosynthesizing plants (NASA Earth Observatory, 2011).

The results also showed the ability of vegetation to recover after a drought period such as an increase in NDVI in the year 2012 after the 2011 east Africa drought which may be attributed to a wet period in the late 2011 and early 2012 (AghaKouchak 2015).

Figure 4.1 shows a relatively dry area indicating sparse vegetation. Mean NDVI ranged between 0.0 to 0.6 indicating a region with shrubs and grasslands (Ndungu et. al., 2019). However, there were periods with healthy vegetation, with NDVI values over 0.5 such as in the years 2007, where Kenya experienced heavy downpours at the beginning of October 2006 (International Federation of Red Cross and Red Crescent Societies, 2008), El-Nino events in 2009/2010 (GoK, 2013) and a wet period in 2012 (AghaKouchak 2015). The year

2005 was marked with the lowest NDVI indicating an extreme dry period due to failure of the October – December short rainy season (Hastenrath et. al., 2007; Wakabi, 2006) resulting in poor vegetation condition.

Figure 4.2 indicates a region with relatively healthy vegetation, with mean NDVI ranging between 0.3 to 0.8, indicating a region with dense vegetation (Ndungu et. al., 2019). NDVI values were over 0.5 for most of the study period.

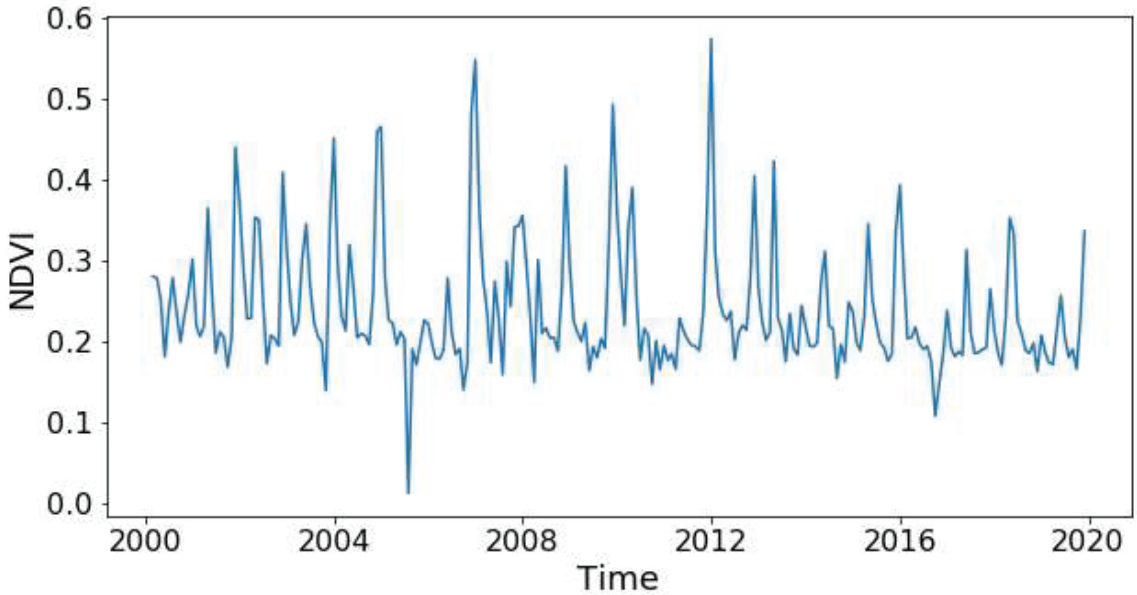

Figure 4.1 . 1 × 1 pixel NDVI time series in the low vegetation region from February 2000 to September 2019. The pixel was extracted from the top right corner of the h21v09 MODIS tile.

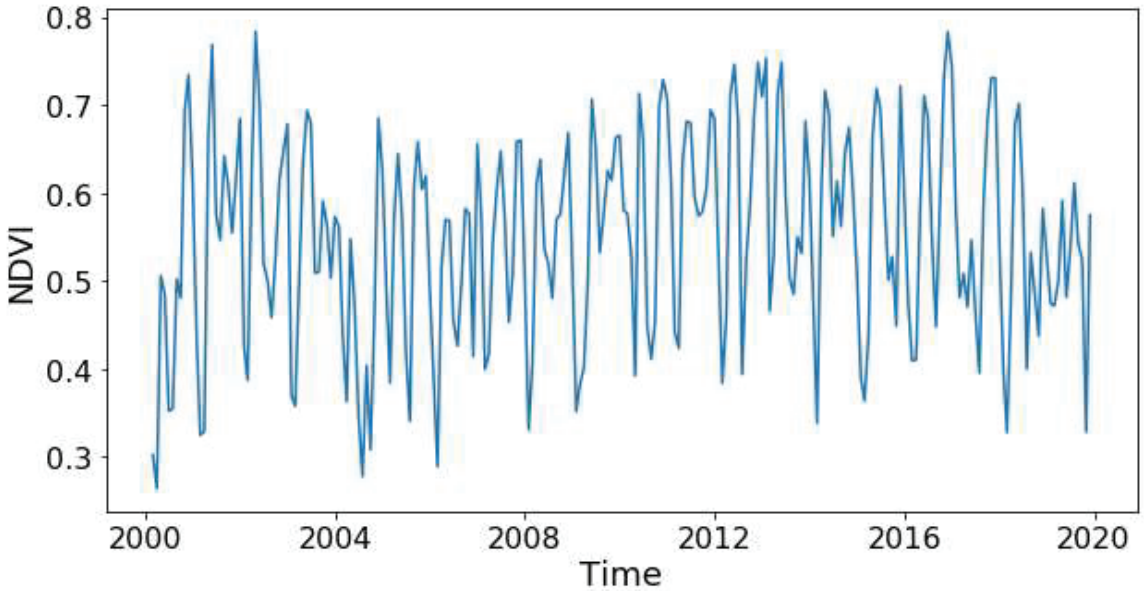

Figure 4.2. 1 × 1 pixel NDVI time series in the high vegetation region from February 2000 to September 2019. The pixel was extracted from the bottom left corner of the h21v08 MODIS tile.

4.3 Analysis of results using the Holt-Winters model

The section below provides the temporal, spatial and statistical analysis of the results using the Holt-Winters model in the low and high vegetation regions.

4.3.1 Observed and predicted NDVI time series (1 × 1 pixel) using the Holt-Winters model in the low and high vegetation regions

Figures 4.3 and 4.4 presents the 1 × 1 pixel time series for the training, testing and forecasted NDVI for in the low and high vegetation regions respectively. The training data was plotted using a blue line while the red and green lines indicates the testing and forecasted data respectively. In both regions, the time series for predicted NDVI using the Holt-Winters model underestimated as well as overestimated during the study period. This may be attributed to climatic conditions such as temperature and rainfall which may cause unusual NDVI values which in turn may cause distortion to the optimized values of the smoothing parameters thereby causing estimation errors in the Holt-Winters model (Goodwin, 2010).

Furthermore, a high NDVI value in the current month may cause an overestimation, thus, the next month's prediction will also be high.

In Figure 4.3, the Holt-Winters model forecasted low NDVI values between 0.2 - 0.4, indicating a region with sparse vegetation. The Holt-Winters model underestimated as well as overestimated most of the study period. However, it follows a similar pattern as the observed NDVI at the end of the study period. In the high vegetation region (Figure 4.4), the Holt-Winters model presented a good model as the NDVI values ranged between 0.5 - 0.7, denoting a region with good and healthy vegetation. The Holt-Winters model overestimated, especially towards the end of the study period, with an exception in 2018, where it followed similar behaviour as the NDVI data. The outliers may be due to the vegetation's phenological cycles as a result of seasonal changes or human activities among other factors (Wang et. al., 2016).

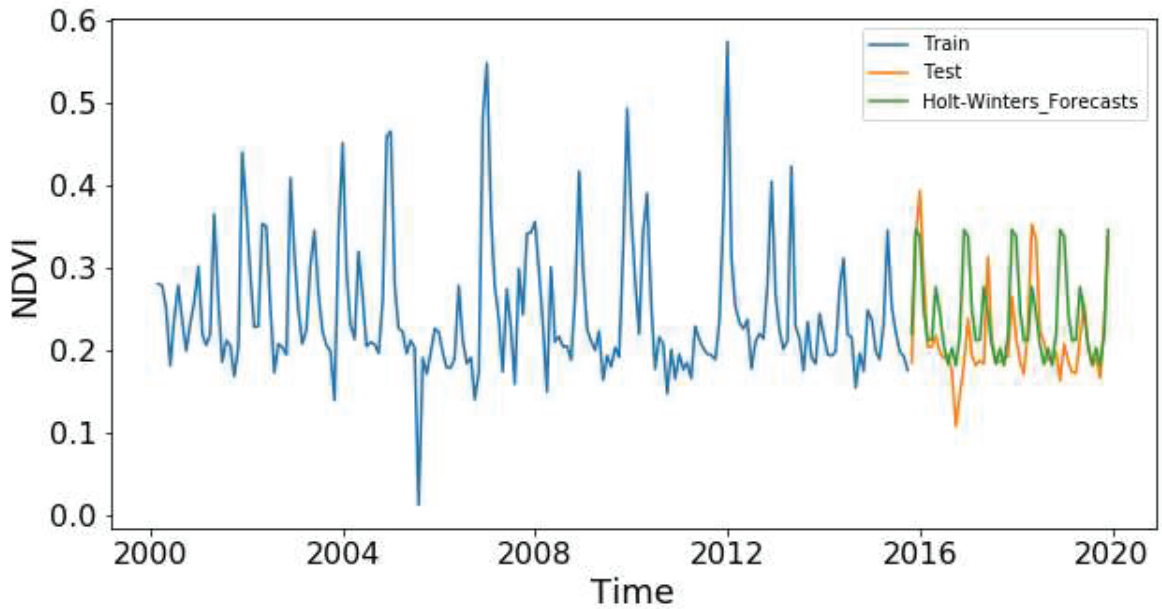

Figure 4.3. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data using Holt-Winters model (green) in the low vegetation region.

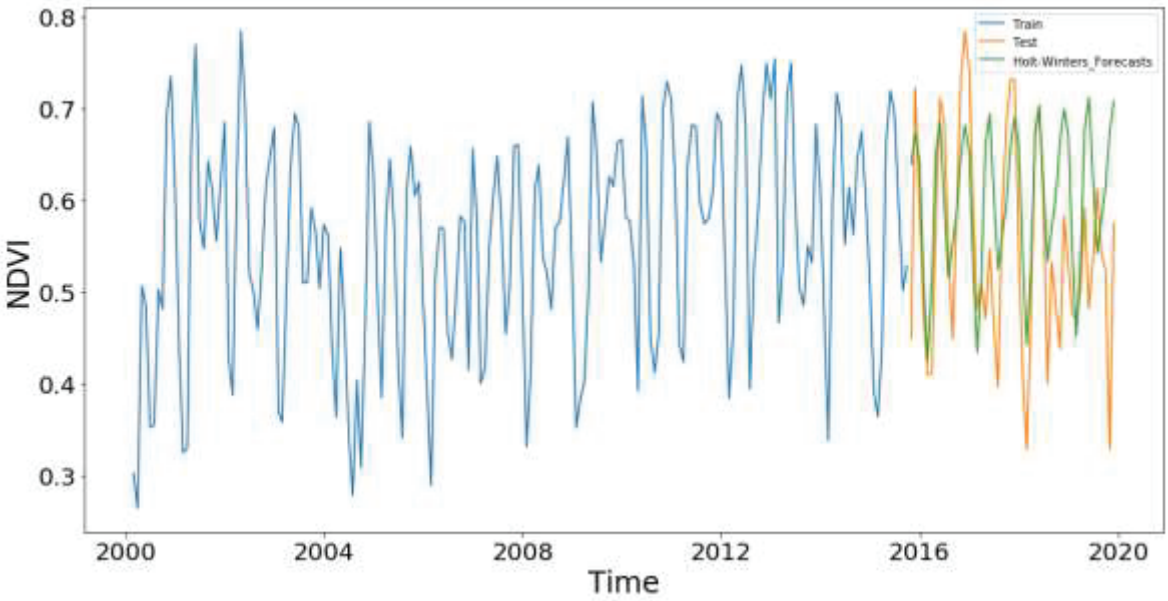

Figure 4. 4. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data (green) using Holt-Winters model in the high vegetation region

4.3.2 Spatial analysis for the 6 months NDVI forecast using the Holt-Winters model in the low and high vegetation regions

The prediction results for the 6 months lead NDVI forecast in CSV format was converted into spatial maps to assess the vegetation condition at the pixel level for the low and high vegetation regions. The resulting maps visualized NDVI for 12 categories (Figure 4.5 and Figure 4.6), with the lowest category indicating a water body and the highest category indicating extremely high vegetation condition. Figure 4.5 and 4.6 shows the spatial maps for the 6 months forecast NDVI in the low and high vegetation regions respectively.

A large section of the low vegetation region (Figure 4.5) was predicted to have moderate to light vegetation deficit in the first month (October 2019), with areas such as Garissa and parts of Tana River counties (Figure 3.1), in the severe vegetation deficit category. This can be attributed to the climate conditions in the previous month as sunny and dry weather in September 2019 experienced in ASAL regions led to low NDVI in October 2019. The above normal rainfall during the October – December 2019 season (NDMA, 2019) led to improved vegetation condition in November and December 2019, with most of the

region predicted to experience moderate vegetation deficit conditions as compared to the previous month, with light vegetation observed in Kitui county (Figure 3.1) and along the river bank.

The spatial distribution of vegetation in January 2020 was predicted to reduce compared to December 2019, with almost the whole region experiencing light vegetation deficit, with the exception of areas along the river bank which was predicted to have moderate vegetation. This may be as a result of the onset of the dry season and an increase in temperature. Persistent poor climatic conditions in February 2020, may have caused a severe vegetation deficit mostly in Garissa county, with a high spatial distribution in the moderate vegetation deficit category. The vegetation condition reduced further in March 2020 as low NDVI was predicted in the entire region as a result of prevailing dry conditions during the dry season (January – March 2020).

The high vegetation region (Figure 4.6) was predicted to experience moderate to extremely high vegetation condition. This may be due to the above normal rainfall experienced in many districts in Uganda and the DRC (Mayuge District Local Government 2020; FEWS NET 2020) may have contributed to good vegetation condition in the high vegetation region for the 6 months (October 2019 to March 2020) as shown in Figure 4.6.

4.3.3 Statistical analysis of the Holt-Winters' performance in the low and high vegetation regions

To quantify the performance of the Holt-Winters model in the low and high vegetation regions, the MAE and RMSE values were computed. Tables 4.1 – 4.3 presents the MAE and RMSE values for 10 × 10, 100 × 100 and 600 × 600 pixels respectively. Forecast error measures for the 600 × 600 pixels were also calculated using MAPE as showed in Table 4.4.

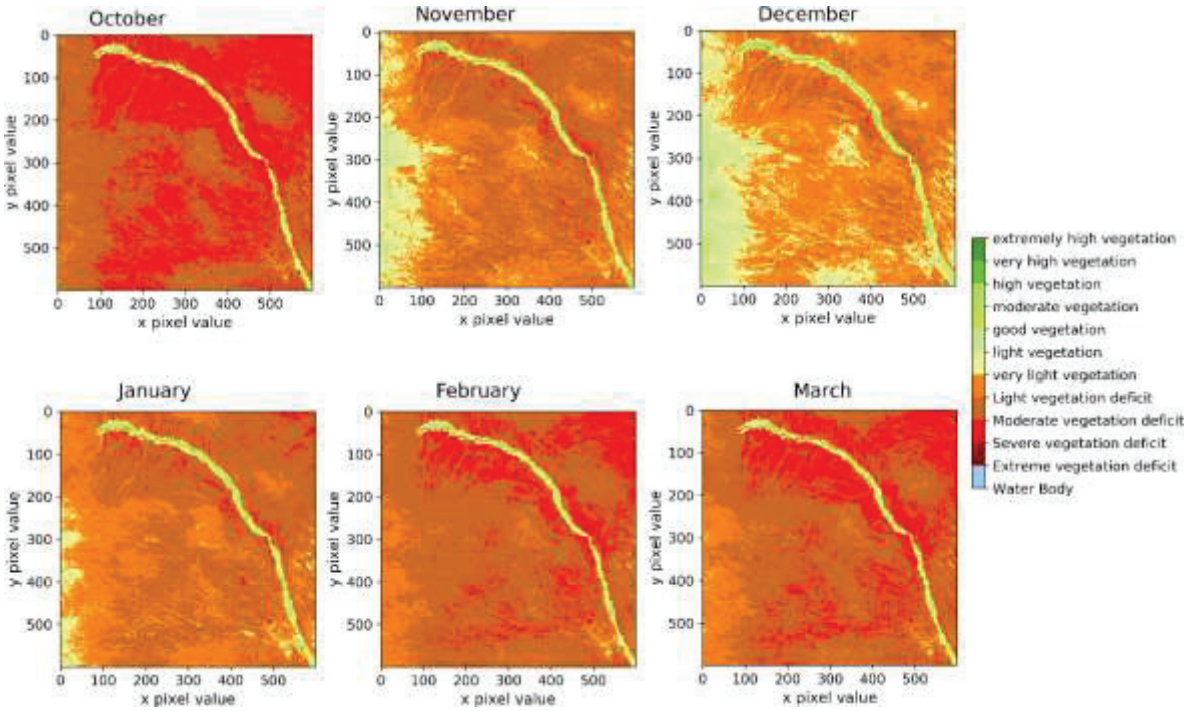

Figure 4. 5. Spatial distribution for the 6 months NDVI forecast using Holt-Winter model in the low vegetation region

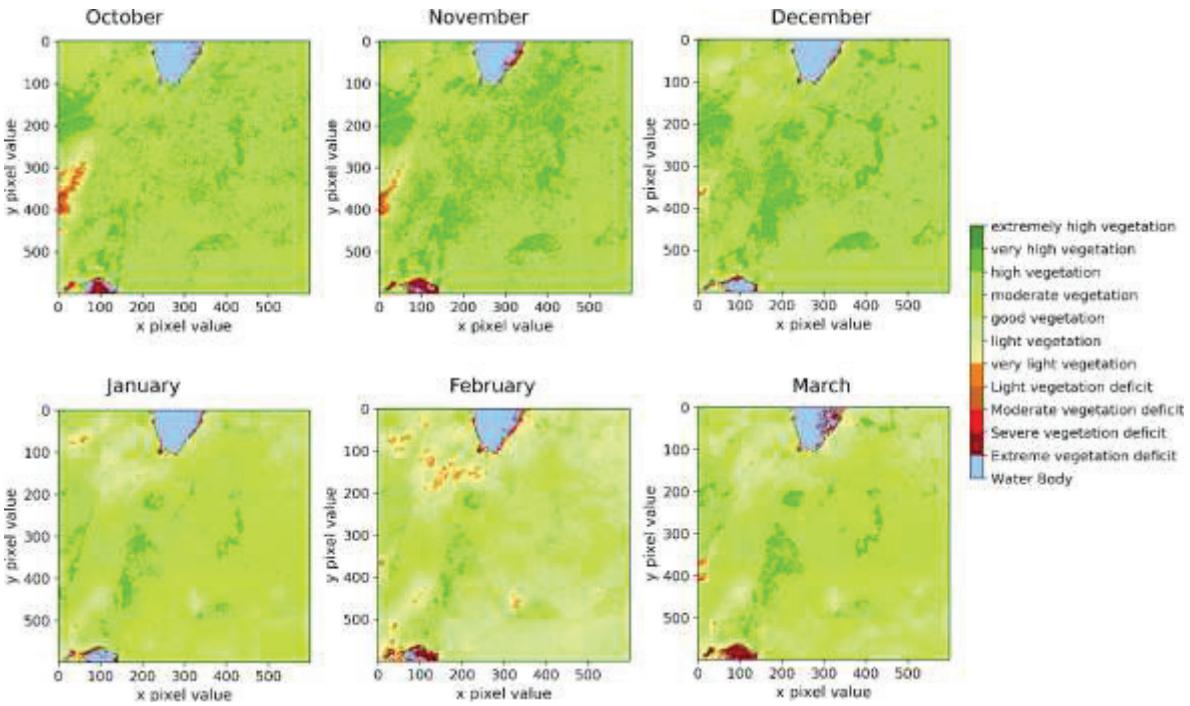

Figure 4. 6.Spatial distribution for the 6 months NDVI forecast using Holt-Winters model in the high vegetation region.

Overall, the model performed better in the low vegetation region compared to the high vegetation region for the 10 × 10 and 100 × 100 pixels, with the lowest MAE and RMSE of 0.0389 and 0.0519 for 10 × 10 pixels and lowest MAE and RMSE of 0.0392 and 0.0565 for 100 × 100 pixels respectively. Although NDVI is criticized of saturation problem in high biomass conditions (Gao et. al., 2000; Tucker 1977), the Holt-Winters model performed better in the high vegetation region compared to the low vegetation region for the 600 × 600 pixels achieving the lowest MAE and RMSE of 0.0679 and 0.084 respectively. This result might be caused by the weak reception ability of the NDVI, due to the reception of the green light spectrum from the high vegetation region. If the density of leaves and plants are dense, the green light reflection from leaves and plants can be easily received while the reflection from land surface or soil surface is blocked by the receiver of imagery satellite photo-sensors. On the other hand, in the low vegetation region, the strong light reflection from land or soil surface, which passes through plants and leaves are not blocked by them, causing huge estimation errors in vegetation estimation. That is why there was poor estimation results in the low vegetation region for the 600 × 600 pixels. The result may also be attributed to NDVI's calculation which is sensitive to atmosphere, soil and pixel component and the spectral response to these factors is not exactly the same in the two spectral bands used for NDVI calculation (Huang et. al., 2021; Julien, 2008).

The validation results between the forecasted and actual data for the low and high vegetation regions are presented in Table 4.4. In the high vegetation region, MAPE resulted in an infinite value as a result of division by zero (Puah et. al., 2016). MAE and RMSE resulted in finite values and were used to overcome the problem of MAPE. The results from the forecast error measures showed that the Holt-Winters model achieved significant accuracy based on the RMSE values as the model showed satisfactory accuracy and is good at predicting NDVI by learning seasonal changes. The forecast error measures also showed that the Holt-Winters model achieved better accuracy in the high vegetation region than in the low vegetation region for 600 × 600 pixels.

Table 4.1. Evaluation of the Holt-Winters additive model in the low and high vegetation regions for 10 × 10 pixels. (All computations were done on a remote server, instance Shape: VM.Standard2.8).

Region	MAE	RMSE	Computation time (sec)
High vegetation	0.0488	0.0615	14.1140
Low vegetation	0.0389	0.0519	19.9854

Table 4. 2. Evaluation of the Holt-Winters additive model in the low and high vegetation regions for 100 × 100 pixels. (All computations were done on a remote server, instance Shape: VM.Standard2.8).

Region	MAE	RMSE	Computation time (sec)
High vegetation	0.0629	0.0787	3186.0698
Low vegetation	0.0392	0.0565	1239.9131

Table 4.3. Evaluation of the Holt-Winters additive model in the low and high vegetation regions for 600 × 600 pixels. (All computations were done on a remote server, instance Shape: VM.Standard2.8).

Region	MAE	RMSE	MSE	Computation time (sec)
High vegetation	0.0679	0.084	0.0071	38847.0363
Low vegetation	0.0744	0.096	0.0092	38378.3088

Table 4. 4. Forecast error measures between the actual and predicted NDVI for the low and high vegetation regions.

Region	MAE	RMSE	MAPE
High vegetation	0.0906	0.1253	∞
Low vegetation	0.1186	0.1468	28.1183

4.4 Analysis of results using the MLP models

The section below provides the temporal, spatial and statistical analysis of the results using the MLP models in the low and high vegetation regions.

4.4.1 Observed and predicted NDVI time series (1 × 1 pixel) using the MLP models in the low and high vegetation regions

Figures 4.7 to 4.9 and Figures 4.10 to 4.12 presents the 1 × 1 pixel time series for the training, testing and forecasted NDVI using the MLP models for in the low and high vegetation regions respectively. The training data was plotted using a blue line while the red and green lines indicates the testing and forecasted data respectively. In both regions, the time series for the forecasted NDVI using the MLP models underestimated as well as overestimated during the study period. As seen in all figures for both the low and high vegetation regions, the MLP 1 and MLP 3 models overestimated NDVI as a result of overfitting due to the ReLU AF (Nwankpa et. al., 2018). Moreover, the ReLU AF is fragile during training causing some of the gradients to die (Nwankpa et. al., 2018) which leads to some neurons being dead leading to zero activation and hindering learning, thus, causing the weights not to activate in future points (Goodfellow et. al., 2016).

However, the MLP 2 model underestimated NDVI but followed a similar pattern as the test data at the end of the study period for the low vegetation region. On the other hand, the MLP 2 underestimated as well as overestimated NDVI for the high vegetation region.

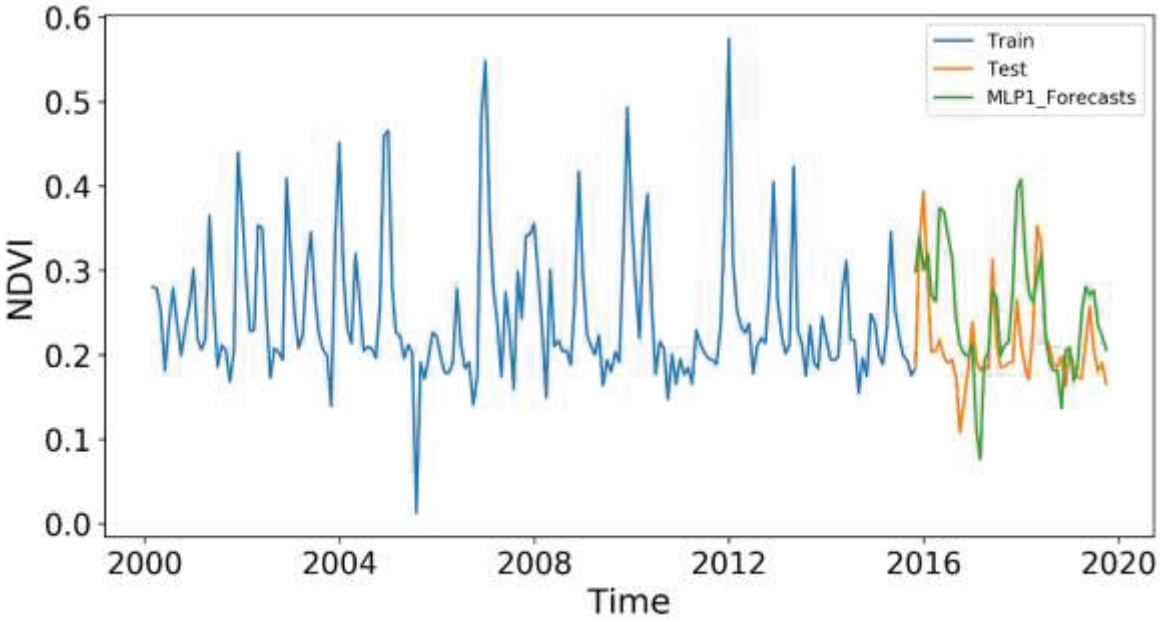

Figure 4. 7. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data (green) using MLP 1 model in the low vegetation region

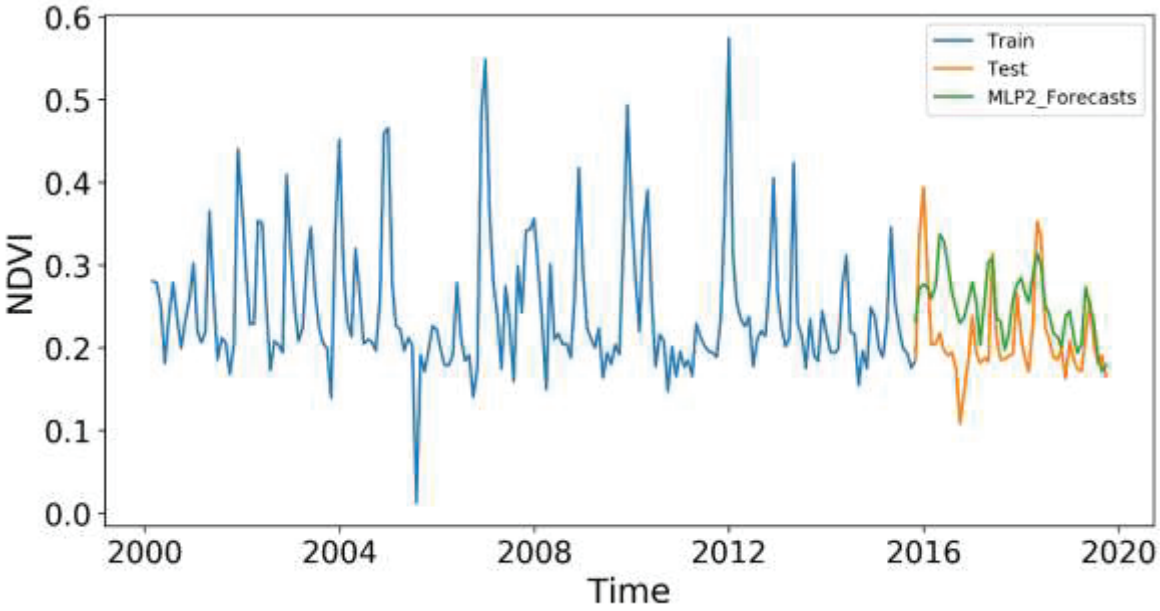

Figure 4. 8. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data (green) using MLP 2 model in the low vegetation region

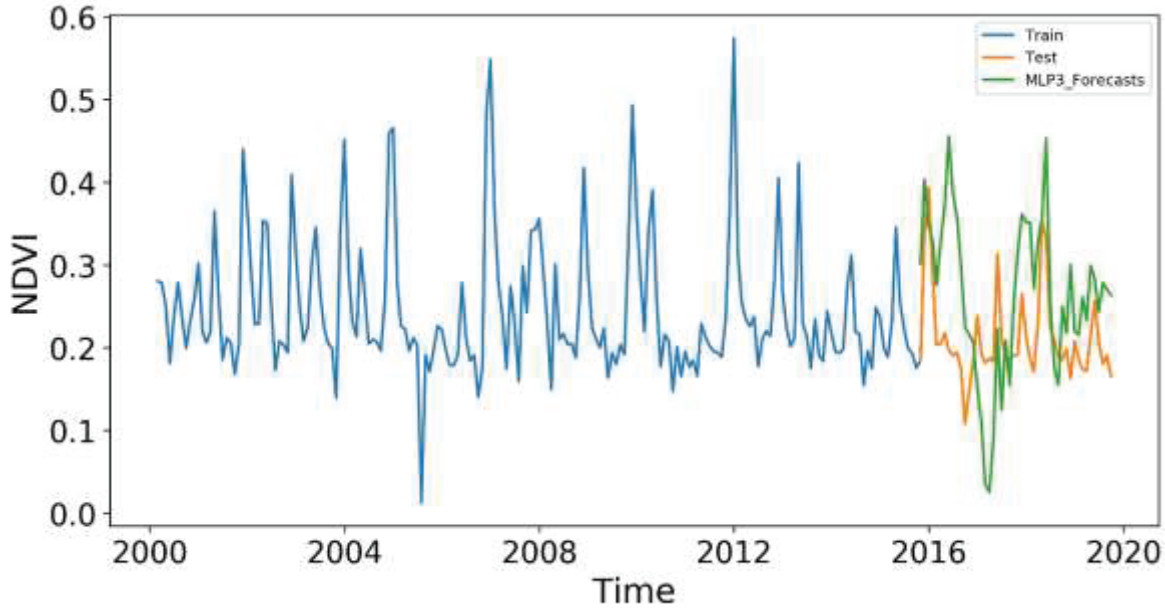

Figure 4. 9. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data (green) using MLP 3 model in the low vegetation region

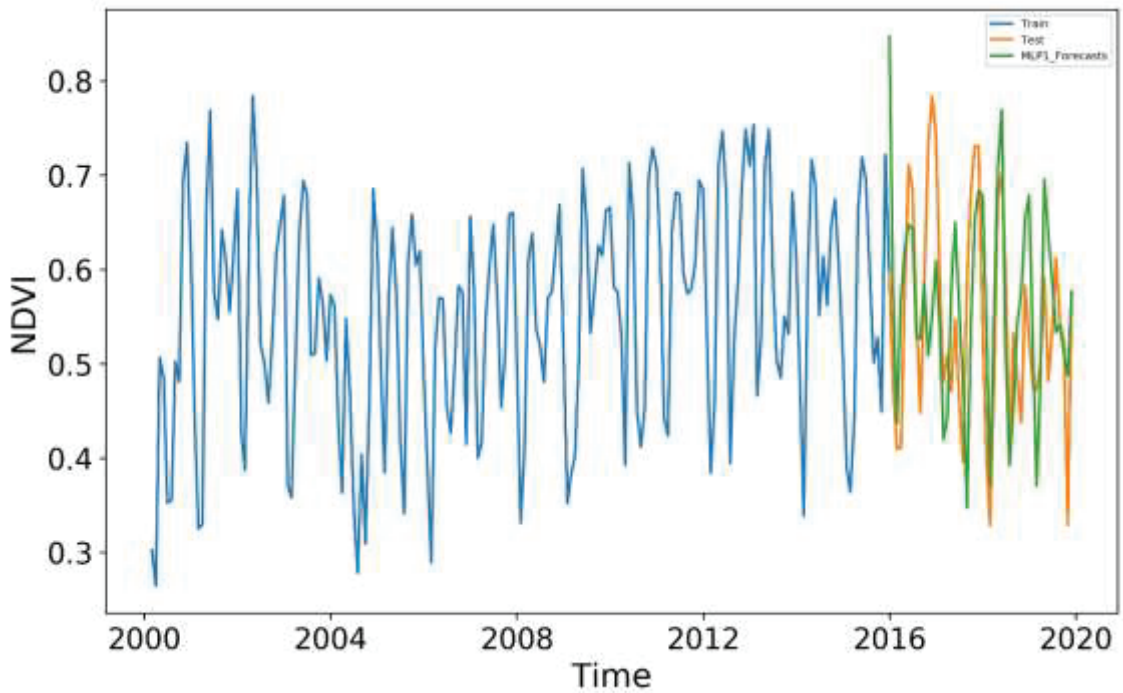

Figure 4. 10. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data (green) using MLP 1 model in the high vegetation region

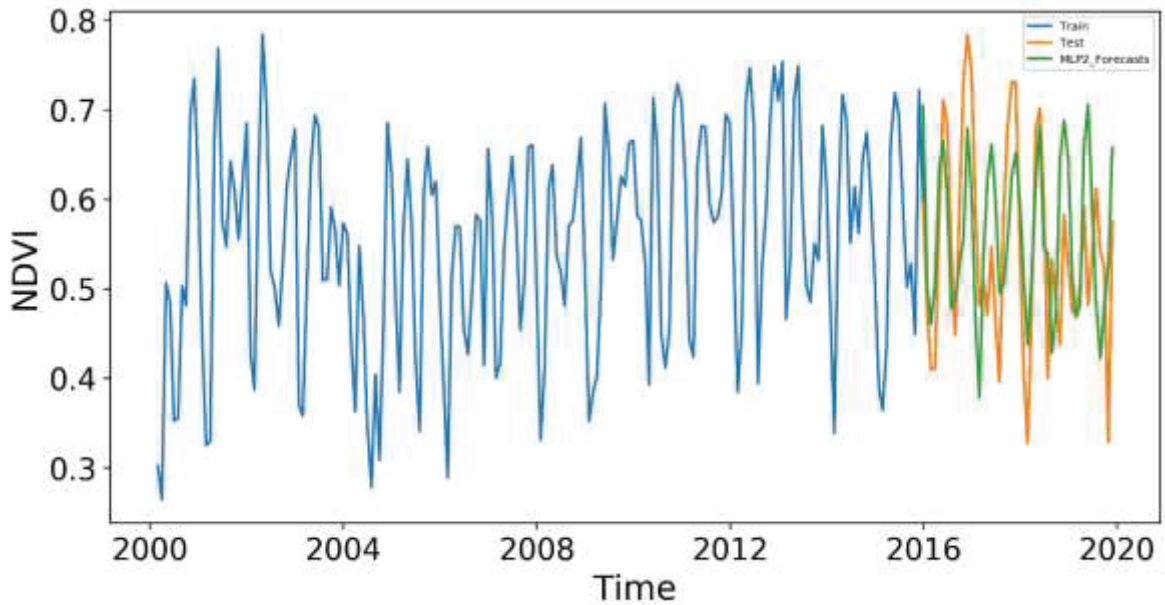

Figure 4. 11. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data (green) using MLP 2 model in the high vegetation region

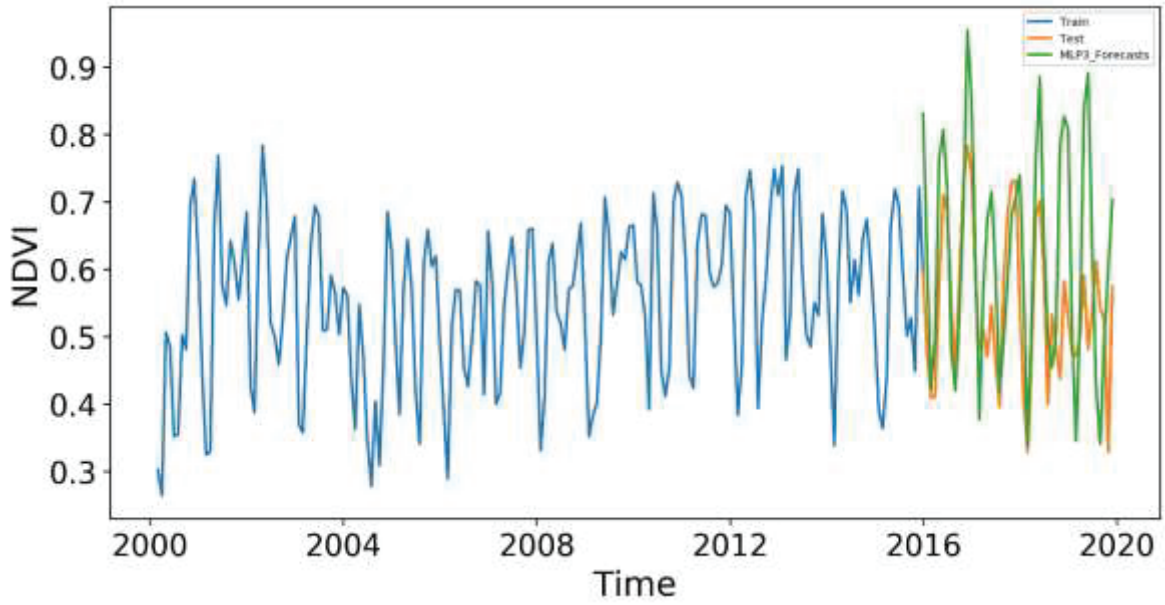

Figure 4. 12. 1 × 1 pixel NDVI time series of the training data (blue), testing data (red) and forecasted data (green) using MLP 3 model in the high vegetation region

4.4.2 Spatial analysis for the 6 months NDVI forecast using the MLP model in the low and high vegetation regions

The three MLP models were used to forecast NDVI during the short rain season (October – December 2019) and during the dry season (January – March 2020) in the low and high vegetation regions. The time-lead prediction was achieved using the previous month's data to forecast NDVI of the following month, such that, using September 2019 data to forecast NDVI for October 2019, until all 6 months lead predictions was completed. Overall, the MLP models can be used for multi-step lead forecasting of NDVI using univariate pixel-wise NDVI data without additional data such as precipitation fed into the neural networks.

Similarly, to the Holt-Winters model, the prediction results using the MLP models for the 6 months lead NDVI forecast in CSV format was converted into spatial maps to assess the vegetation condition at the pixel level for the low and high vegetation regions. The resulting maps visualized NDVI for 12 classes: between -1 to +1 (Figure 4.13 to Figure 4.18). As seen in the figures, the lowest class (-1) indicates a water body, while the highest class (+1) indicates high NDVI value, which inturn means extremely good vegetation condition. The forecast maps for the two seasons are presented in Figures 4.13 to 4.15 for MLP 1, MLP 2 and MLP 3 respectively for the low vegetation region and Figures 4.16 to 4.18 for MLP 1, MLP 2 and MLP 3 models respectively for the high vegetation region.

The forecast maps for the two seasons are presented in Figures 4.13 to 4.15 for MLP 1, MLP 2 and MLP 3 respectively for the low vegetation region. The forecasted NDVI maps for the MLP 1 (Figure 4.13) showed that the whole study area was predicted to have low NDVI in the range between 0.0 to 0.2 during the short rain season (October – December 2019) and dry season (January – March 2020). Few pixels on the west side of the region have green patches indicating moderate NDVI, while the region along the river bank was predicted to have higher NDVI values (above 0.5) due to the thriving of vegetation throughout the year.

The forecasted NDVI maps from the MLP 3 (4.15) model showed a similar behavior as the forecasted maps from the MLP 1 model. On the other hand, the forecasted NDVI maps from the MLP 2 (Figure 4.14) showed moderate NDVI (between 0.2 – 0.5) for a large section

of the study area, while some pixels in a few spots have low NDVI of 0.1. Similarly, to the forecasted maps from MLP 1 and MLP 3, the NDVI along the river bank is higher due to the thriving vegetation. Although the low vegetation region is an ASAL, vegetation thrives along the river bank due to the ability of the river bed to retain water (The European Space Agency, 2021). Generally, the NDVI in the entire study area was predicted to range between 0.2 to 0.6 during the October – December 2019 period which may be attributed to the short rains season, however, the NDVI was forecasted to be lower in the range between 0.0 to 0.4 during the January – March 2020 period which may be a result of the dry season and an increase in temperature.

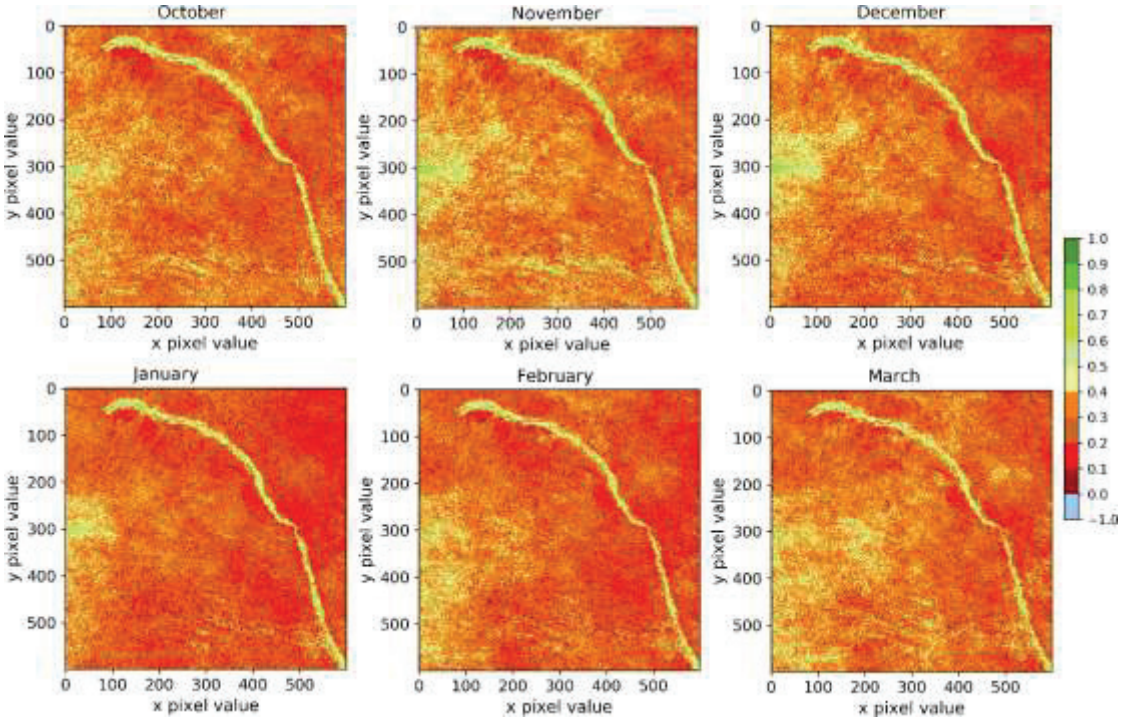

Figure 4. 13. Spatial distribution for the 6 months lead NDVI forecast maps produced by MLP 1 in the low vegetation region.

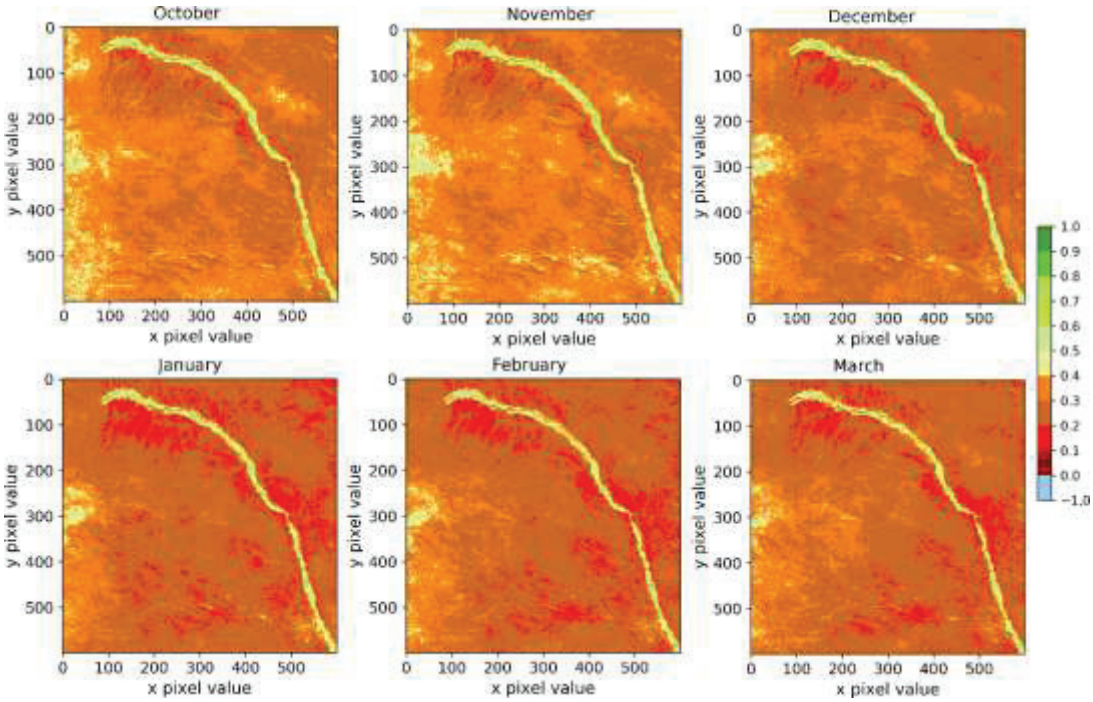

Figure 4. 14. Spatial distribution for the 6 months lead NDVI forecast maps produced by MLP 2 in the low vegetation region

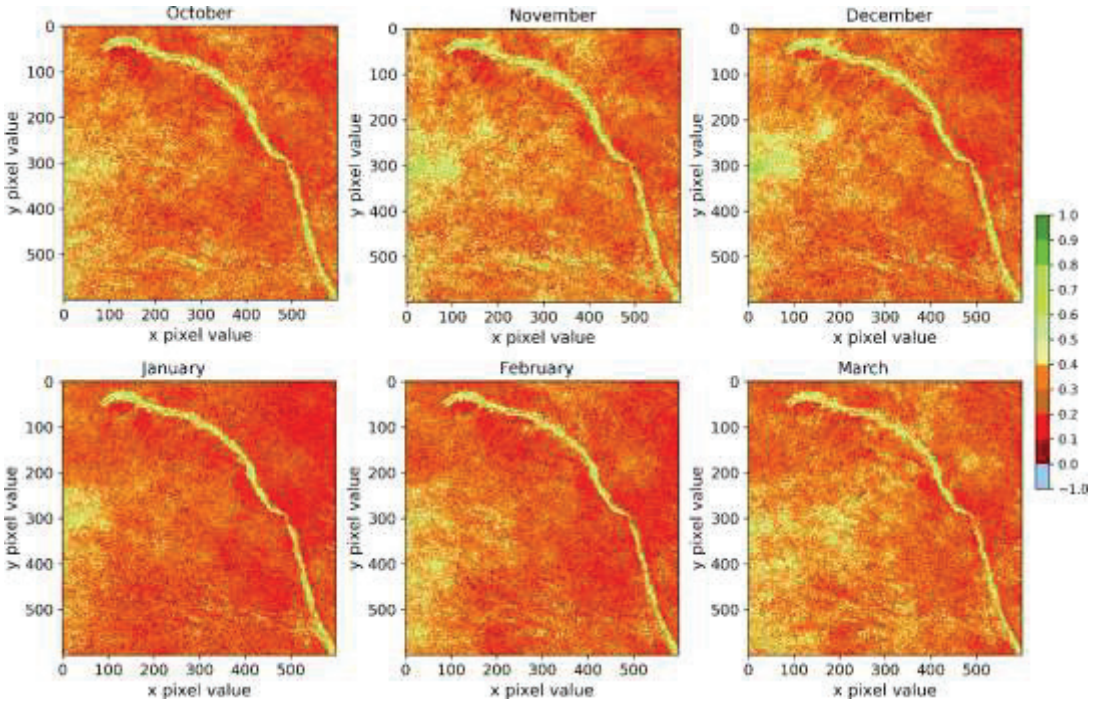

Figure 4. 15. Spatial distribution for the 6 months lead NDVI forecast maps produced by MLP 3 in the low vegetation region

The spatial distribution for the 6 months lead NDVI forecast maps for the high vegetation region are presented in Figures 4.16 to 4.18 for MLP 1, MLP 2 and MLP 3 models respectively. Generally, all the models predicted high NDVI values ranging between 0.5 to 1.0 in the region during the October – December 2019 and January – March 2020 seasons. This may be due to the above normal rainfall experienced in many districts of Uganda and the DRC (Mayuge District Local Government 2020; FEWS NET 2020) which may have contributed to good vegetation condition in the region.

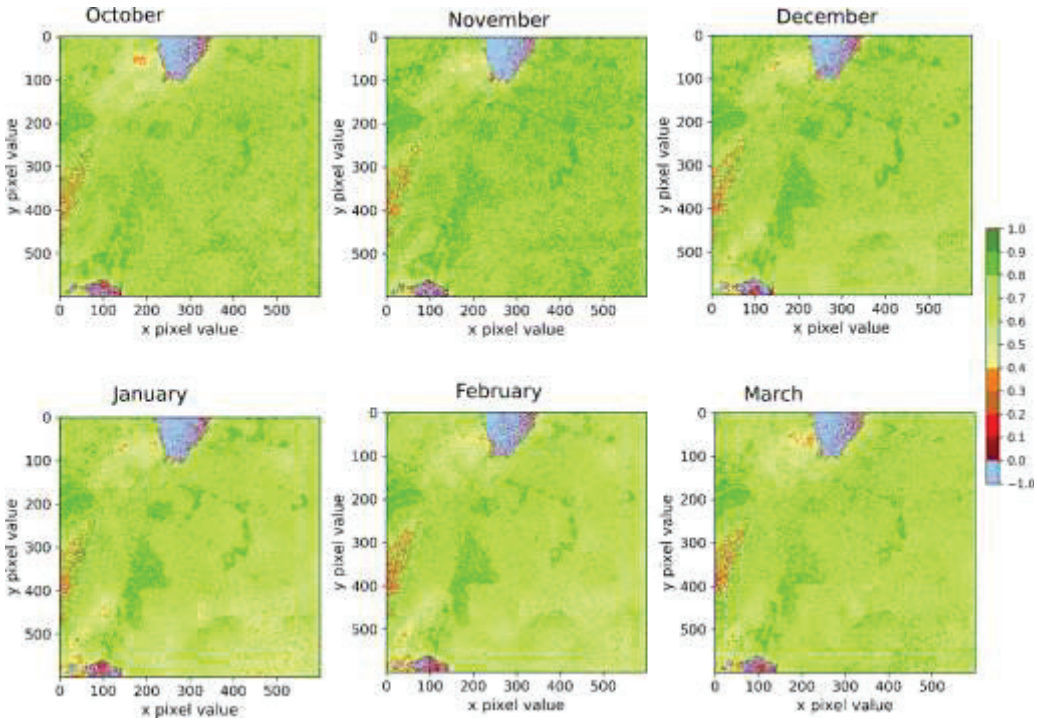

Figure 4. 16. Spatial distribution for the 6 months lead NDVI forecast maps produced by MLP 1 in the high vegetation region

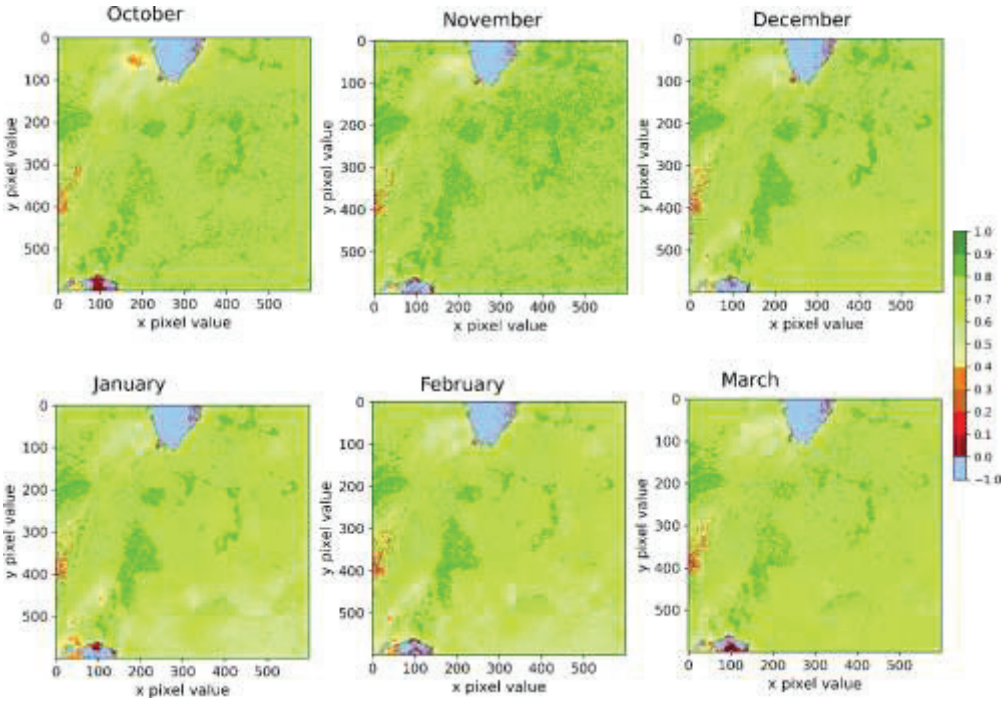

Figure 4. 17. Spatial distribution for the 6 months lead NDVI forecast maps produced by MLP 2 in the high vegetation region

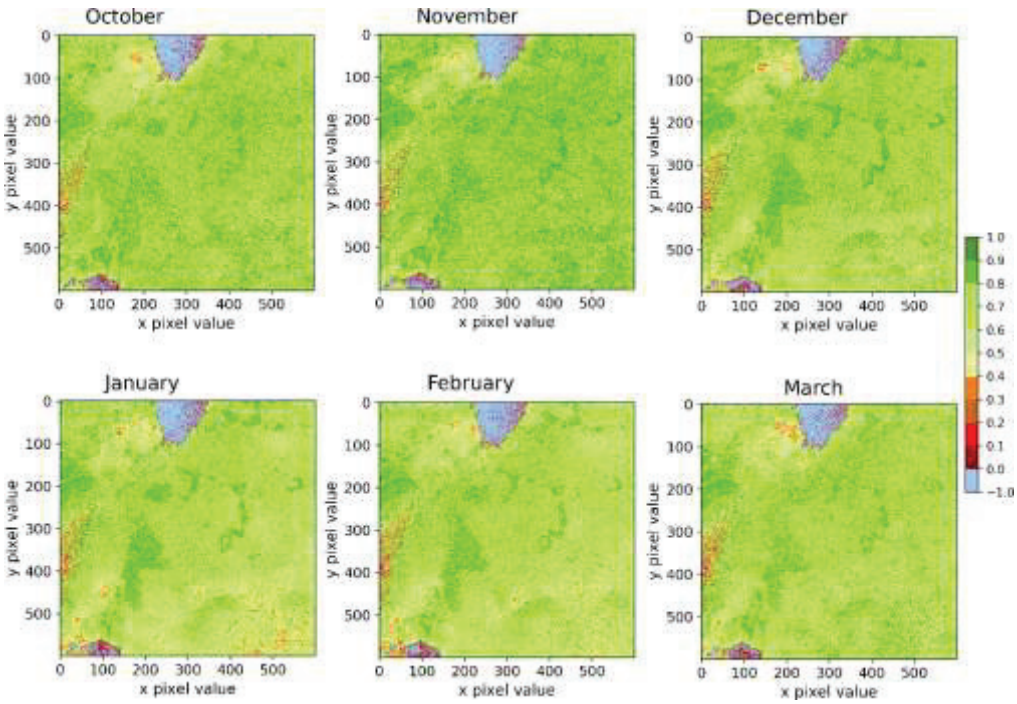

Figure 4. 18. Spatial distribution for the 6 months lead NDVI forecast maps produced by MLP 3 in the high vegetation region

4.4.3 Statistical analysis of the MLP models in the low and high vegetation regions

Tables 4.5 and 4.6 shows the statistical analysis for the MLP models in the low and high vegetation regions respectively. The MLP 2 model achieved better predictive performance in both regions with the lowest RMSE, MSE and MAE. The MLP 2 showed better predictive performance as TanH AF provides better training performance for multi-layer neural networks (Neal, 1992) and produces zero centered output thus, aiding in the back propagation process (Nwankpa et. al., 2018). In addition, the network training converges faster if the average of each input variable is close to zero, resulting in a minimal cost function. On the other hand, the MLP 3 model produced the worst performance in both regions achieving the lowest RMSE, MSE and MAE.

Table 4. 5. Evaluation of MLP models in the low vegetation region for 600 × 600 pixels. (All computations were done on the supercomputer)

Model	RMSE	MSE	MAE
MLP 1 (3000 epochs, ReLU, Adam)	0.11180	0.01603	0.09643
MLP 2 (4000 epochs, TanH, Adamax)	0.07207	0.00589	0.06417
MLP 3 (5000 epochs, ReLU, Nadam)	0.11918	0.01797	0.10179

Table 4. 6. Evaluation of MLP models in the high vegetation region for 600 × 600 pixels. (All computations were done on the supercomputer)

Model	RMSE	MSE	MAE
MLP 1 (3000 epochs, ReLU, Adam)	0.10840	0.01590	0.09073
MLP 2 (4000 epochs, TanH, Adamax)	0.08479	0.00935	0.07189
MLP 3 (5000 epochs, ReLU, Nadam)	0.12569	0.02054	0.02054

4.5 Analysis of results using the SARIMA model

The section below provides the temporal and statistical analysis of the results using the SARIMA model in the low and high vegetation regions. Due to budget and time constraints and limitation of computational resources, it is worth mentioning that the SARIMA model for 600 × 600 pixels was done for the low vegetation region only. However, the statistical analysis was performed for 100 × 100 pixels for both the low and high vegetation regions.

4.5.1 Observed and predicted NDVI time series (1 × 1 pixel) using SARIMA model in the low and high vegetation regions

Figures 4.19 and 4.20 show the 1 ×1 NDVI time series for the training, testing and predicted values using the SARIMA model in the low and high vegetation regions. In Figure 4.19, the observed NDVI ranged between 0.0 to 0.6 during the study period, with the lowest observed NDVI value at 0.0 in the year 2005 indicating that the region experienced scarce vegetation condition during that year which can be attributed to a prolonged dry period. On the other hand, the observed NDVI ranged between 0.3 to 0.8 in the high vegetation region as shown in Figure 4.20. NDVI values were above 0.5 most of the years during the study period.

As shown in Figures 4.19 and 4.20, the SARIMA model underestimated at the peaks and overestimated at the troughs preventing it from fitting to the testing data, which led to a poor forecast. In addition, the SARIMA model predicted that the values further ahead in time will be even lower. Thus, despite the model serving as a general estimation of the NDVI time series trend, it is insufficient and leads to inaccurate forecasts. Exogenous data such as rainfall may be used to capture the heterogeneous rainfall distribution which may improve the performance of the SARIMA model (Mutti et. al., 2020).

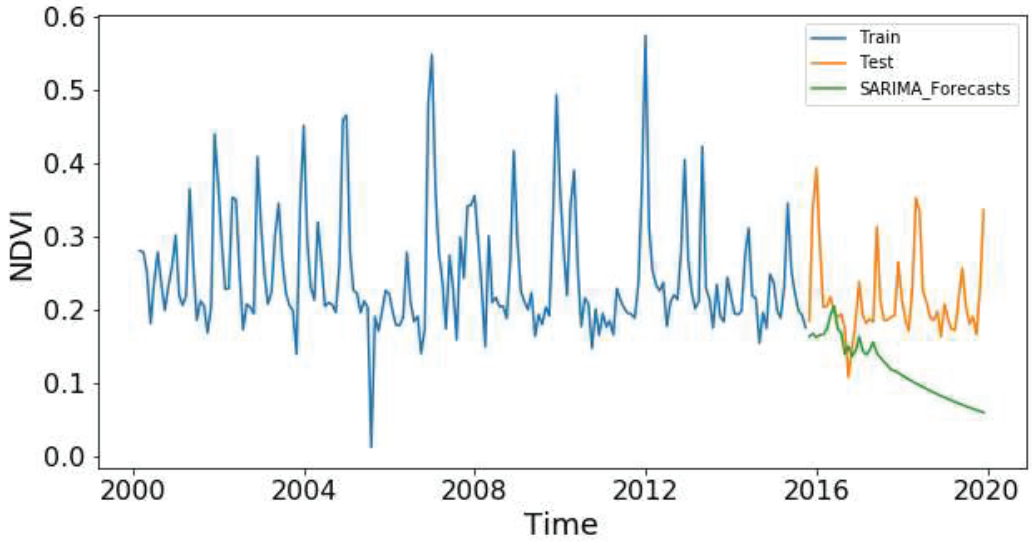

Figure 4. 19. 1 ×1 pixel monthly mean NDVI time series, showing the training data (blue), testing data (red) and the predicted data using SARIMA model (green) in the low vegetation region.

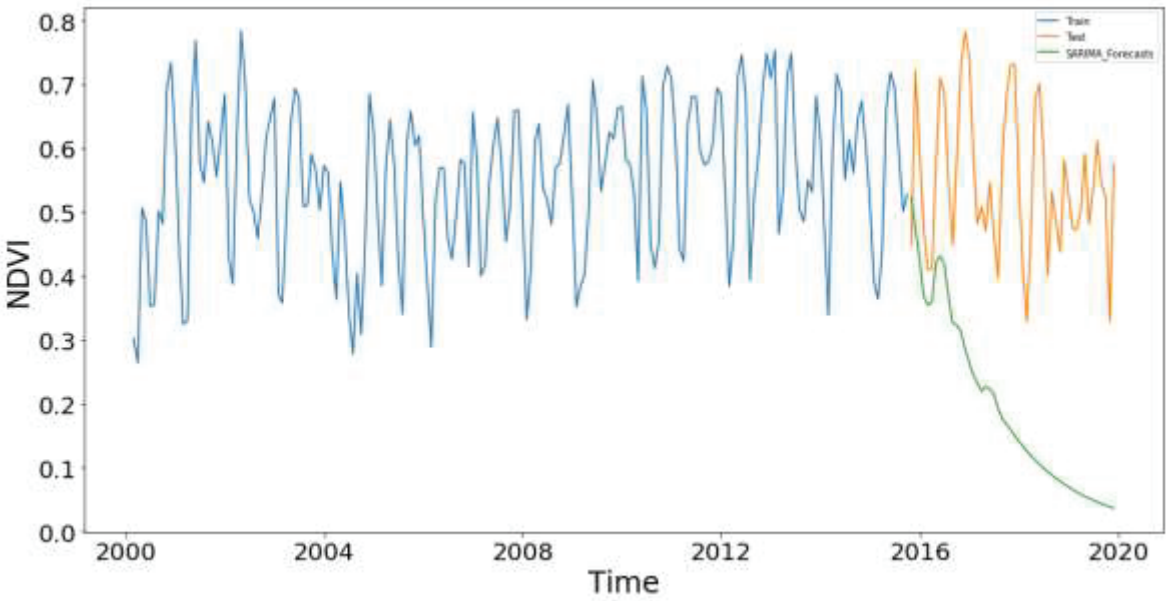

Figure 4. 20. 1 × 1 pixels monthly mean NDVI time series, showing the training data (blue), testing data (red) and the predicted data using SARIMA model (green) in the high vegetation region.

4.5.2 Statistical analysis of the SARIMA model in the low and high vegetation regions using 100 ×100 pixels

Due to budget and time constraints and limitation of computational resources, the SARIMA model for the high vegetation region was analyzed using 100 × 100 pixels instead of the 600 × 600 pixels. Table 4.7 shows the statistical analysis for the SARIMA model in the low and high vegetation regions respectively using 100 × 100 pixels. Despite the down scaling of the number of pixels to 100 × 100, the assumption is that the results for the 600 × 600 pixels using the SARIMA model for the high vegetation region will be poor. An evaluation of the results from the temporal analysis of the SARIMA model with the other models (Holt-Winters and MLP) using the 1 × 1 pixel showed that the SARIMA model does not fit well to the test data, thus leads to poor predictions, with a downward trend for future predictions. Thus, the SARIMA model presents a poor estimation of NDVI. Moreover, comparing the results from 100 × 100 pixels using the Holt-Winters model (Table 4.2) showed that the SARIMA model achieved the worst predictions for the high vegetation region. According to Brito et. al., (2021), SARIMA models are more reliable when forecasting longer intervals (two years ahead) and their limitations after being trained on short observation periods are due to overfitting problems. This may explain why the SARIMA model had the worst potential for the estimation of NDVI in the high vegetation region since forecasting in this research was for mid-term (6 months) only.

Table 4. 7. Evaluation of the SARIMA model in the low and high vegetation regions for 100 × 100 pixels. (All computations were done on the supercomputer)

Region	RMSE	MSE	MAE
Low vegetation	0.1235	0.0165	0.1052
High vegetation	0.1318	0. 0246	0.108

4.6 Comparative analysis of the models in forecasting NDVI in the low and high vegetation regions

This section presents the comparative analysis of the models (Holt-Winters, SARIMA and MLP) in the forecasting of NDVI in the low and high vegetation regions. The proposed models in this research were validated to assess how well the models forecasted NDVI using statistical indicators described in Chapter 3, section 3.6. Table 4.8 and 4.9 presents the results of the performance of the models in the low and high vegetation regions respectively. The best models were the MLP 2 model (4000 epochs, TanH activation function and Adamax optimizer) and the Holt-Winters model for the low and high vegetation regions respectively. These models outperformed the other models in all the three statistical metrics, indicating that the models performed well with the data. In both models (MLP 2 and the Holt-Winters), the MAE (0.06417 and 0.0679 respectively) is significantly lower than the RMSE (0.07207 and 0.084 respectively) indicating a small variance in the error.

As shown in Table 4.8, the MLP 2 model (4000 epochs, TanH activation function and Adamax optimizer) was more accurate and had better metrics than the other models using univariate pixel-wise NDVI time series data. The RMSE, MSE and MAE (0.07207, 0.00589 and 0.06417 respectively) is significantly smaller for the MLP 2 compared to the other models, indicating more accurate forecasting ability. On the other hand, the SARIMA model was the worst performing model having the worst metrics, thus may lead to inaccurate results.

Table 4.9 shows that the Holt-Winters model achieved better predictive performance in the high vegetation region with the lowest RMSE, MSE and MAE of 0.084, 0.0071 and 0.0679 respectively. On the other hand, the MLP 3 model (5000 epochs, ReLU, Nadam) was the worst performing model with the lowest RMSE, MSE and MAE of 0.12569, 0.02054 and 0.10476 respectively.

Table 4.8. Evaluation of the models using statistical metrics for 600 × 600 pixels in the low vegetation region

Model	RMSE	MSE	MAE
Holt-Winters (additive)	0.096	0.0092	0.0744
SARIMA(1, 0, 1)(0, 0, 2)$_{12}$	0.1344	0.0181	0.1112
MLP 1 (3000 epochs, ReLU, Adam)	0.11180	0.01603	0.09643
MLP 2 **(4000 epochs, TanH, Adamax)**	**0.07207**	**0.00589**	**0.06417**
MLP 3 (5000 epochs, ReLU, Nadam)	0.11918	0.01797	0.10179

Table 4. 9. Evaluation of the models using statistical metrics for 600 × 600 pixels in the high vegetation region

Model	RMSE	MSE	MAE
Holt-Winters (additive)	**0.084**	**0.0071**	**0.0679**
MLP 1 (3000 epochs, ReLU, Adam)	0.10840	0.01590	0.09073
MLP 2 (4000 epochs, TanH, Adamax)	0.08479	0.00935	0.07189
MLP 3 (5000 epochs, ReLU, Nadam)	0.12569	0.02054	0.10476

The results from this research suggests that using univariate pixel-wise NDVI data, the MLP 2 and Holt-Winters models (for the low and high vegetation regions respectively) can provide better results for RMSE, MSE and MAE, hence, it is concluded that the two models performed well for the two regions in the study area. Overall, the NDVI forecasting models in this research produced satisfactory results and allowed statistical indicators such as RMSE, MSE and MAE to be used to evaluate the models. The outcomes from the models showed good performance as values < 0.5 suggests good performance of prediction models (Can et. al., 2005).

CHAPTER FIVE

CONCLUSION

This chapter concludes with a summary of the research findings, limitations encountered when carrying out the research, some recommendations for policy makers and future work.

The possibilities of using remote sensing data and NDVI in particular, for monitoring and forecasting vegetation conditions are great. The east Africa region mainly engage in agriculture and livestock keeping and is vulnerable to natural disasters such as droughts, famine and floods. In addition, the region experiences increased vegetation degradation mainly caused by human activities and climate changes such as increasing temperature and inadequate rainfall, especially in the ASAL regions.

Normalized difference vegetation index (NDVI) can be used to differentiate vegetation cover across a region by revealing vegetation changes and identify regions of vegetation deficit using temporal and spatial analysis of NDVI time series data. This can be achieved using NDVI time series plots to analyze temporal patterns and maps to analyze the spatial distribution to determine the overall vegetation condition in a region.

This research assessed and forecasted vegetation condition using univariate pixel-wise NDVI data (from February 2000 – September 2019) in east Africa in a low and high vegetation regions. MODIS NDVI data was used at it gives a continuous time series with a global spatial coverage and short temporal duration. In addition, the moderate resolution data is helpful to capture field scale variability.

Pixel-wise modeling provides robustness, although it requires higher computational power, thus, creates a need to use high speed computing. The most crucial step during the models' development was to select the best hyperparameters as they are the key to develop forecasting models which will result in accurate predictions. Thus, to minimize the computational time required to develop the models using the entire dataset for the study area, a subset of the data (1 $\times$ 1 pixel) was used to analyze the temporal pattern of the NDVI time series, grid search the models' hyperparameters and eventually develop the models.

Temporal analysis of NDVI is useful to monitor and detect vegetation condition as it provides information on vegetation phenological phases by low or high NDVI values. The

low and high vegetation regions have different climatic and vegetation condition. The temporal pattern of NDVI showed that the vegetation in the regions had undergone significant changes since 2000 to 2019. NDVI had a distinct temporal pattern as observed from the 1 × 1 pixel NDVI time series analysis, with the high vegetation region having high NDVI values and the low vegetation region having low NDVI values. Both regions had periods of high and low NDVI values such as a low NDVI in 2011, followed by a high NDVI in 2012.

The comparison of the forecasting performance of the Holt-Winters model in the low and high vegetation regions was conducted at three levels: 10 × 10, 100 × 100 and 600 × 600 pixels which is helpful to guide similar research based on computational capacity of institutions, especially in developing countries. However, this research placed more importance on the results from the 600 × 600 pixels as one of the objectives was to use a large dataset in order to cover a larger area which may be applied to forecast vegetation condition in other regions with similar characteristics. Thus, the results from the higher pixel level were used as they covered a larger area as well as give an insight on the expected results for similar research in the east Africa region.

NDVI is often criticized of saturation problem in high biomass regions, however, the result obtained in this research showed that the Holt-Winters model achieved better predictive performance in the high vegetation region compared to the low vegetation region for the 600 × 600 pixels with the lowest MAE, RMSE and MSE of 0.0679, 0.084 and 0.0071 respectively. The better predictive performance of the model in the high vegetation region may be due to the reception of the green light spectrum which causes the green light to be easily received by satellite photo sensors whereas for the low vegetation region, the strong light reflection from the land or soil surface are not blocked resulting in poor prediction.

The comparative analysis between the models showed that the MLP 2 model (4000 epochs, TanH AF and Adamax optimizer) showed the best skill to forecast NDVI in the low vegetation region at 6 months' lead time, while the SARIMA model showed the worst skill. The MLP 2 using the TanH activation function showed better predictive performance as TanH activation function produces zero centered output thus, aiding in the back propagation

process. Moreover, the network training converges faster if the average of each input variable is close to zero, resulting in a minimal cost function. ANN models have been widely accepted as useful methods for modeling complex non-linear relationships in time series data. Thus, the good performance of the MLP 2 model in this research has provided new evidence of such properties of these models. On the other hand, the Holt-Winters model achieved the best predictive performance while the MLP 3 model resulted in the worst performance in the high vegetation region.

Overall, the models showed reasonable predictive skills in forecasting NDVI at 6 months' lead time and may be helpful in forecasting NDVI in order to reduce vegetation degradation as well as for decision making in ecosystem and biodiversity conservation. Moreover, the 6 months' lead time NDVI forecast will assist forest managers and land planners to effectively plan and devise methods that will ensure proper utilization of vegetation to help minimize and overcome vegetation degradation.

Spatial mapping of NDVI could reveal thriving vegetation or vegetation deficit regions which may be studied to determine the factors affecting vegetation condition such as climate change, anthropogenic activities or plant' phenological changes. Thus, the forecasted NDVI spatial maps revealed clear patterns of regions with thriving or deficit vegetation which may suggest that there are other variables which may influence vegetation condition. The maps revealed that although the low vegetation region is an ASAL, the region along the river bank has high NDVI values.

Although NDVI is affected by other factors such as rainfall, temperature, soil moisture etc., this research demonstrated that it is possible to monitor and forecast vegetation condition in a region using NDVI data only as NDVI can be forecasted by using its lagged values only.

Even though this research proposed a novel method in forecasting NDVI using pixel-wise modeling, it is however, as with the majority of researches, subject to several limitations. The NDVI data used had a comparatively shorter temporal length compared to NOAA AVHRR NDVI data, (which has more than 30 years' data, though with a lower spatial resolution of 0.05º, approximately 5.5 km) which may influence the effectiveness of the

forecasting models used in this research. Thus, temporal analysis using long term NDVI historical databases could help characterize vegetation condition and the implications humans have on it at the local and global scale.

As noted before, vegetation condition may be affected by climatic conditions such as rainfall and temperature. However, existing rainfall and temperature data are in coarse spatial resolution such as 27.25 km resolution for rainfall data from Tropical Rainfall Measuring Mission (TRMM), 55 km resolution rainfall data from Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) and 1 km resolution land surface temperature (LST) data from MODIS Terra Land Surface Temperature/Emissivity Daily (MOD11A1) or MODIS Terra Land Surface Temperature/Emissivity 8-Day (MOD11A2). Due to the coarse spatial resolution of such data, it became difficult to perform a correlation analysis between rainfall, temperature and NDVI using the pixel-wise approach. In addition, the coarse spatial resolution of the rainfall and temperature data made it difficult to have a correct coordinate match with the moderate resolution NDVI data used in this research for pixel-wise modeling.

Accurate and timely data that describes vegetation condition is crucial to assess vegetation development and to best understand and monitor land use practices. The challenge of meeting local and global demand in land usage while conserving and protecting vegetation will require different stakeholders to focus on mitigation of vegetation degradation through the use of readily and freely available remote sensing data and developing forecasting tools to allow making informed decisions that are environmentally friendly and economically feasible.

The forecasting models used in this research should be used with caution since their behavior will change and adapt differently depending on the NDVI data used and the characteristics of a region such as climatic conditions, types of vegetation, land use and land cover etc.

This study was limited to two regions (low and high vegetation regions) in east Africa. The NDVI forecasts spatial maps at the pixel level revealed areas with vegetation deficit which may offer opportunities for further research to focus on how local communities are

adapting to the impacts of vegetation degradation and the mitigation methods they are using to reduce the impacts.

Although in-situ data provides better opportunity to identify actual cases of vegetation degradation, the data relies on traditional methods of data collection which are time consuming and costly. Thus, policy makers can utilize various data sources such as the freely available remote sensing data and various assessment tools.

The results of this research may promote increased awareness for the local communities, especially in the low vegetation region to use vegetation and land resources sustainably and avoid practices that promote vegetation degradation for long-term use of the resources.

Future research perspectives will be to explore other ANN models such as convolutional neural networks (CNN) as well as perform the analysis for short-term, mid-term and long-term forecasts.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the Almighty God for His mercy, good health and all the blessings throughout my life and during my study at the University of Hyogo.

I would like to express my sincere gratitude and appreciation to my PhD advisor and mentor, Professor Hajime Kawamukai, for his encouragement, advice, valuable guidance and suggestions throughout the research and development of this thesis. His kind patience, and financial support led to the success of this research. I would also like to express special thanks to Professor Tomoe Entani and Professor Ryo Haraguchi for their insightful comments and valuable feedback in improving my thesis.

I would like to thank Kawanishi Memorial Shin Meiwa Education Foundation for granting me financial support for 1 year which helped in my academic and research work.

Many thanks to the faculty members and staff of the Graduate School of Applied Informatics, University of Hyogo for their assistance during my study.

I am grateful to the Land Processes Distributed Active Archive Center (LPDAAC) for providing the MOD13Q1 MODIS/Terra Vegetation Indices data and the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) for providing the shapefiles used to create the maps.

This research partly used computational resources of Research Institute for Information Technology, Kyushu University. I appreciate Professor Teruo Tanimoto of the Research Institute for Information Technology, Kyushu University for his assistance in setting up the supercomputer for my usage.

Finally, I would like to express my deepest gratitude to my family for their love, support and encouragement. I am forever indebted to my parents for their selfless sacrifices during my upbringing and for giving me an education. My heartfelt thanks and appreciation goes to my husband: Islam, for his understanding, patience, tolerance and support to accomplish this research and my kids: Fatma and Hassan for their patience and sacrifices which enabled me to complete this research.

REFERENCES

Abadi, M., Agarwal, A., Barham, P. et. al. (2015) TensorFlow: large-scale Machine learning on heterogeneous systems. Available from: https://www.tensorflow.org/

Abera T. A. (2020) Climatic impacts of vegetation dynamics in Eastern Africa. PhD. Dissertation, University of Helsinki, Unigrafia, Helsinki.

Abraham A. (2005) 129 Artificial neural networks in Sydenham P. H. & Thorn R. Handbook of Measuring System Design.

Adler R. F., Gu G., Sapiano M., Wang J. -J., Huffman J. (2017) Global precipitation: Means, variations and trends during the satellite era (1979 – 2014). Surveys in Geophysics 38, pp. 679 – 699.

African Union (2015) Agenda 2063: Report of the commission on the African Union Agenda 2063. Available from: https://au.int/agenda2063

AghaKouchak A. (2015) A multivariate approach for persistence-based drought prediction: Application to the 2010-2011 east Africa drought. Journal of Hydrology 526, pp. 127 – 135.

Agwata J. F. (2005) Water resources utilization, conflicts and interventions in the Tana Basin in Kenya. Proceedings, Topics of Integrated Watershed Management, FWU, vol. 3. pp. 13 – 23.

Ahrends A., Burgess N. D., Milledge S. A. H., Bulling M. T., Fisher B., Smart J. C., Clarke G. P., Mhoro B. E., Lewis S. L. (2010) Predictable waves of sequential forest degradation and biodiversity loss spreading from an African city. Proceedings of the National Academy of Sciences USA 107: 14556 – 14561.

Alhamad M. N., Stuth J., Vannucci M. (2007) Biophysical modelling and NDVI time series to project near-term forage supply: Spectral analysis aided by wavelet denoising and ARIMA modelling. International Journal of Remote Sensing 28 (11), pp. 2513 – 2548.

Allen D. M. (1971) Mean square error of prediction as a criterion of selecting variables. Technometrics 13 (3), pp. 469 – 475.

Anyamba A., Tucker C. J. (2005) Analysis of Sahelian vegetation dynamics using NOAA-AVHRR NDVI data from 1981-2003. Journal of Arid Environments 63 (3), pp. 596 – 614.

Aryee G., Essuman R., Djagbletey R., Darkwa E. O. (2019) Comparing the forecasting performance of seasonal Arima and Holt-Winters methods of births at a tertiary healthcare facility in Ghana. Journal of Biostatistics and Epidemiology 5 (1), pp. 18 – 27.

Balkin S. D., Ord J. K. (2000) Automatic neural network modelling for univariate time series. International Journal of Forecasting 16, pp. 509 – 515.

Baker T., Kiptala J., Olaka L., Oates N., Hussein A., McCartney M. (2015) Baseline Review and Ecosystem Services Assessment of the Tana River Basin, Kenya. International Water Management Institute. Colombo, Sri Lanka. IWMI Working Papers. 165.

Barney B. (2020) Introduction to parallel computing. Lawrence Livermore National Laboratory. Available from: https://computing.llnl.gov/tutorials/parallel_comp/

Belayneh A., Adamowski J., Khalil B., Quilty J. (2016). Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. Atmospheric Research 172 – 173, pp. 37 – 47.

Bélair C., Ichikawa K., Wong B. Y. L., Mulongoy K. J. (eds). (2010) Sustainable use of biological diversity in socio-ecological production landscapes. Background to the 'Satoyama Initiative for the benefit of biodiversity and human well-being', Secretariat of the Convention on Biological Diversity, Montreal, CBD Technical Series no. 52, p. 184.

Bello I., Zoph B., Vasudevan V., Le Q. V. (2017) Neural optimizer search with reinforcement learning. Proceedings of the 34th International Conference on Machine Learning.

Bernard K. N. (1998) State of forest genetic resources in Kenya, Sub-Regional Workshop FAO/IPGRI/ICRAF on the conservation, management, sustainable utilization and enhancement of forest genetic resources in Sahelian and North-Sudanian Africa, Forest

Genetic Resources Working Papers, Working Paper FGR/18E, Forestry Department, FAO. Rome, Italy.

Bergstra J., Bengio Y. (2012) Random search for hyper-parameter optimization. Journal of Machine Learning Research 13, pp. 281 – 305.

Benardos P. G., Vosniakos G. - C. (2007) Optimizing feedforward artificial neural network architecture. Engineering Applications of Artificial Intelligence 20 (3), pp. 365 – 382.

Beukering P., Moel H. (eds)., et. al. (2015) The economics of ecosystem services of the Tana River Basin, assessment of the impact of large infrastructural interventions. Institute of Environmental Studies, University of Amsterdam. Report number: R15-03.

Bhandari S. P. (2011) Monitoring forest dynamics using time series of satellite image data in Queensland, Australia. PhD. Dissertation, The University of Queensland. p. 180.

Bishop C. M. (1995) Neural networks for pattern recognition. Oxford University Press, Oxford, UK.

Bosch B. (ed), Douma P., Bolhuis E., Klaver D., Zawadi Y. (2009) Public private cooperation fragile states, field study report DR Congo, as part of the country study DR Congo. Report of field research mission to north Kivu and Ituri.

Box G. E. P., Jenkins G. M. (1970) Time series analysis: Forecasting and control. San Franscisco, California, Holden-Day.

Bradley B. A., Mustard J. F. (2008) Comparison of phenology trends by land cover class: A case study in the Great Basin, USA. Global Change Biology 14, pp. 334 – 346.

Brink A. B., Eva H. D. (2009) Monitoring 25 years of land cover change dynamics in Africa: A sample based remote sensing approach. Applied Geography 29 (4), pp. 501 – 512.

Brito G. R. A., Villaverde A. R., Quan A. L., Pérez E. R. (2021) Comparison between SARIMA and Holt-Winters models for forecasting monthly streamflow in the western region of Cuba. SN Applied Sciences 3, 671

Brown R. G. (1959) Statistical Forecasting for inventory control. Mc Graw-Hill (Inc.), New York, USA.

Brownlee J. (2020) What is time series forecasting? Available from: https://machinelearningmastery.com/time-series-forecasting/

Brownlee J. (2020) How to grid search triple exponential smoothing for time series forecasting in python. Available from: https://machinelearningmastery.com/how-to-grid-search-triple-exponential-smoothing-for-time-series-forecasting-in-python/

Brownlee J. (2021) Gradient descent optimization with Nadam from scratch. Available from: https://machinelearningmastery.com/gradient-descent-optimization-with-nadam-from-scratch/

Brownlee J. (2019) A gentle introduction to the rectified linear unit (ReLU). Available from: https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/

Brownlee J. (2018) Difference between a batch and an epoch in a neural network. Available from: https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/

Byrnes R. M. (1992) Uganda: A country study. Washington, D.C.: Federal Research Division, Library of Congress.

Busetto L., Meroni M., Colombo R. (2008). Combining medium and coarse spatial resolution satellite data to improve the estimation of sub-pixel NDVI time series. Remote Sensing of Environment 112 (1), pp. 118 – 131.

Camelo H. N., Lucio P. S., Junior J. B. V. L., de Carvalho P. C. M. dos Santos D. G. (2018). Innovative hybrid models for forecasting time series applied in wind generation based on the combination of time series models with artificial neural networks. Energy 151, pp. 347 – 357.

Campbell J. B. (1987) Introduction to remote sensing. The Guilford Press, New York, USA, 551.

Can T., Nefeslioglu H. A., Gokceoglu C., Sonmez H., Duman T. Y. (2005). Susceptibility assessment of shallow earthflows triggered by heavy rainfall at three sub catchments by logistic regression analyses. Geomorphology 72 (1 – 4), 250 – 271.

Chapman C. A., Chapman L. J. (2004) Unfavorable successional pathways and the conservation value of logged tropical forest. Biodiversity and Conservation 13, pp. 2089-2105.

Charlton M., Caimo A. (2012) Time series analysis. Technical report, European Union, p. 19.

Chatfield C. (1978) The Holt-Winters forecasting procedure. Journal of the Royal Statistical Society: Series C (Applied Statistics), 27 (3), pp. 264 – 279.

Cheng Y., Vrieling A., Fava F., Meroni M, Marshall M., Gachoki S. (2020) Phenology of short cycles in a Kenyan rangeland from PlanetScope and Sentinel-2. Remote Sensing of Environment 248, 112004.

Chicco D. (2017) Ten quick tips for machine learning on computational biology. BioData Mining 10: 35.

Chollet F. (2017). Deep learning with python. Available from: https://keras.io/

Clevert D. -A., Unterthiner T., Hochreiter S. (2016) Fast and accurate deep network learning by exponential linear units (ELU). International Conference on Learning Representations.

County Government of Tana River (2018) Second integrated development plan: Kenya vision 2030, towards a globally competitive and prosperous County, Tana River, Kenya.

Cowpertwait P. S. P., Metcalfe A. V. (2009) Introductory time series with R. Gentleman R., Hornik K., Parnigiani G. (eds). Springer, New York, USA.

Cryer J. D., Chan K-S. (2008) Time series analysis with applications in R. Springer, p. 505.

[Dataset] Didan K. (2015) MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250 m SIN Grid V006. NASA EOSDIS Land Processes DAAC. Available from: https://doi.org/10.5067/MODIS/MOD13Q1.006 . Accessed: 2019 December 10.

Didan K., Munoz A. B., Solano R., Huete A. (2015) MODIS vegetation index user's guide (MOD13 Series), v. 3 (collection 6), Vegetation Index and Phenology Lab, The University of Arizona.

Dogo E. M., Afolabi O. J., Nwulu N. I., Twala B., Aigbavboa C. O. (2018) A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks. Proceedings of the International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS).

Dozat T. (2016) Incorporating Nesterov momentum into Adam. Proceedings of the International Conference on Learning Representations

Du J., Shu J., Yin J., Yuan X., Jiaerheng A., Xiong S., He P., Liu W. (2015) Analysis of spatio-temporal trends and drivers in vegetation growth during recent decades in Xinjiang, China. International Journal of Applied Earth Observation and Geoinformation 38, pp. 216 – 228.

Earth Observing System (EOS). (2020) Types of remote sensing: Technology changing the world. Available from: https://eos.com/blog/types-of-remote-sensing/

Earthdata Search. (2019). Greenbelt, MD. Earth Science Data and Information System (ESDIS) Project. Earth Science Projects Division (ESPD), Flight Projects Directorate. Goddard Space Flight Center (GSFC) National Aeronautics and Space Administration (NASA). Available from: https://earthdata.nasa.gov/gcmd-forum

Eerens H., Haesen D., Rembold F., Urbano F., Tote C., Bydekerke L. (2014). Image time series processing for agriculture monitoring. Environmental Modelling & Software, 53, pp. 154 - 162

e Silva L. P., Xavier A. P. C., da Silva R. M., Santos C. A. G. (2020) Modeling land cover change based on an artificial neural network for a semiarid river basin in northeastern Brazil. Global Ecology and Conservation 21, e00811.

FAO, Food and Agriculture Organization of the United Nations. (1986) Water for animals. Available from: https://www.fao.org/

FAO and UNEP (2020). The state of the world's forests. Forests, biodiversity and people. Rome. Available from: https://doi.org/10.4060/ca8642en

Felix R., Meroni M., Urbano F., Royer A., Atzberger C., Lemoine G., Eerens H., Haesen D. (2015) Remote sensing time series analysis for crop monitoring with the SPIRITS software: New functionalities and use examples. Frontiers in Environmental Science 3: 46, pp. 1 – 11.

Fernández-Manso, A., Quintano C., Fernández-Manso O. (2011). Forecast of NDVI in coniferous areas using temporal ARIMA analysis and climatic data at a regional scale. International Journal of Remote Sensing 32 (6), pp. 1595-1617.

Foley J. A., Levis S., Costa M. H., Cramer W. & Pollard D. (2000) Incorporating dynamic vegetation cover within global climate models. Ecological Applications 10, pp. 1620 – 1632.

Fensholt R., Rasmussen K. (2011) Analysis of trends in the Sahelian 'rain-use efficiency' using GIMMS NDVI, RFE and GPCP rainfall data. Remote Sensing of Environment 115 (2), pp. 438 – 451.

FEWS NET (2020) Democratic Republic of Congo: Food security outlook October 2019 to May 2020. Available from: https://fews.net/sites/default/files/documents/reports/DRC_%20FSO-October%2019_EN_.pdf

FEWS NET (Famine Early Warning Systems Network). (2011) East Africa Food Security Update. Available from: https://fews.net/

Funk C. (2011) We thought trouble was coming. Nature 476 (7)

Fung, K. F., Huang, Y. F., Koo, C. H., Soh Y. W. (2020). Drought forecasting: A review of modelling approaches 2007 – 2017. Journal of Water and Climate Change. 11 (3), pp. 771 – 799.

Gamoun M. (2013) Vegetation change in variable rangeland environments: The relative contribution of drought and soil type in arid rangelands. Ekológia 32 (1), pp. 148 – 157.

Gardner E. S., Jr. (2005) Exponential smoothing: The state of the art – Part II. International Journal of Forecasting 22 (4), pp. 637 – 666.

Gao X., Huete A. R., Ni W., Miura, T. (2000) Optical-biophysical relationships of vegetation spectra without background contamination. Remote Sensing of Environment 74 (3), pp. 609 – 620.

Gascon M., Cirach M., Martínez D., Dadvand P., Valentin A., Plasència A., Nieuwenhuijsen M. J. (2016) Normalized difference vegetation index (NDVI) as a marker of surrounding greenness in epidemiological studies: The case of Barcelona City. Urban Forestry & Urban Greening 19, pp. 88 - 94.

Gelper S., Fried R., Croux C. (2010) Robust forecasting with exponential and Holt-Winters smoothing. Journal of Forecasting 29, pp. 285 – 300.

Gitelson, A. A., (2012). Chapter 15: Remote sensing estimation of crop biophysical characteristics at various scales. in Thenkabail, P. S., Lyon, J. G., Huete, A. (eds), Hyperspectral remote sensing of vegetation. Taylor & Francis Group, New York. pp. 329 –358.

Githui F. W., Mutua F., Bauwens W. (2009) Estimating the impacts of land-cover change on runoff using the soil and water assessment tool (SWAT): Case study of Nzoia catchment, Kenya. Hydrological Sciences Journal 54 (5), pp. 899 – 908.

Glorot X., Bordes A., Bengio Y. (2011) Deep sparse rectifier neural networks. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics.

Goodfellow I., Bengio Y., Courville A. (2016) Deep Learning. MIT Press [Online]. Available from: https://www.deeplearningbook.org/

GoK (Government of Kenya). (2014). Agricultural sector development support programme. Ministry of Agriculture, Livestock and Fisheries, Nairobi, Kenya.

Government of Kenya (GoK), Ministry of Environment & Forestry (2013) El-Nino. Available from: http://www.environment.go.ke/?p=1337

Goodwin P. (2010) The Holt-Winters approach to exponential smoothing: 50 years old and going strong. Foresight: The International Journal of Applied Forecasting, pp. 30 – 33.

Goward S. N., Markhan B., Dye D. G., Dulaney W., Yang J. (1991) Normalized difference vegetation index measurements from the Advanced Very High Resolution Radiometer. Remote Sensing of Environment 35, pp. 257 – 277.

Graupe D. (2013) Principles of Artificial neural networks. 3rd edition. World Scientific Publishing Co. Pte. Ltd. Toh Tuck Link, Singapore.

Gu G., Adler R. F., Huffman J. (2015) Long-term changes/trends in surface temperature and precipitation during the satellite era (1979 – 2012). Climate Dynamics 46, pp. 1091 – 1105.

Guerrier S., Molinari R., Xu H., Zhang Y. (2019) Applied time series with R. (Online book). Available from: https://smac-group.github.io/ts/index.html

Hamid N. A., Nawi N. M., Ghazali R., Salleh M. N. M. (2011) Improvements of back propagation algorithm performance by adaptively changing gain, momentum and learning rate. International Journal on New Computer Architectures and their Applications 1 (3), pp. 889 – 901.

Hansson K., Yella S., Dougherty M., Fleyeh H. (2016). Machine learning algorithms in heavy process manufacturing. American Journal of Intelligent Systems 6 (1), pp. 1 – 13.

Hassoun M. H. (1995) Fundamentals of artificial neural networks. The MIT Press, Cambridge, MA.

Hastenrath S., Polzin D., Mutai C. (2007). Diagnosing the 2005 drought in equatorial east Africa. Journal of Climate 20 (18), pp. 4628-4637.

Hastenrath S., Polzin D., Mutai C. (2011) Circulation mechanisms of Kenya rainfall anomalies. Journal of Climate 24, pp. 404 – 412.

Hatfield J. L., Gitelson A. A., Schepers J. S., Walthall C. L. (2008) Application of spectral remote sensing for agronomic decisions. Agronomy Journal 100 (S3), pp. S-117 – S-131.

Haykin S. (2009) Neural networks and learning machines. 3rd edition. Pearson Education Inc., Upper Saddle River, New Jersey.

Hipel K. W., McLeod A. I. (1994) Time series modelling of water resources and environmental systems. Elsevier Science B. V. Amsterdam, The Netherlands.

Holt C. E. (1957) Forecasting seasonals and trends by exponentially weighted averages (O. N. R. Memorandum No. 52). Carnegie Institute of Technology Pittsburgh, USA.

Hornik K., Stinchcombe M., White H. (1989) Multilayer feedforward networks are universal approximators. Neural Networks 2 (5), pp. 359 – 366.

Huang S., Tang L., Hupy J. P., Wang Y., Shao G. (2021) A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. Journal of Forestry Research 32, pp. 1 – 6.

Huete A., Didan K., Miura T., Rodriguez E. P., Gao X., Ferreira L. G. (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment 83, pp. 195 – 213.

Hunter J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering 9 (3), pp. 90 – 95. Available from: https://matplotlib.org/

Hyndman R. J., Athanasopoulos G. (2018) Forecasting: principles and practice, 2nd edition, OTexts, Melbourne, Australia. Available from: https://otexts.com/fpp2/

Hyndman R., Khandakar Y. (2008) Automatic time series forecasting: The forecast packages for R. Journal of Statistical Software 27 (3), pp. 1 – 22.

Hyndman R. J., Koehler A. B., Ord J. K., Snyder R. D. (2008) Forecasting with exponential smoothing, the state space approach. Springer, Verlag Berlin Heidelberg.

Institute and Faculty of Actuaries (1997). Section B. Description of stochastic models. Claims Reserving Manuals, v.2. Available from: https://www.actuaries.org.uk

International Federation of Red Cross and Red Crescent Societies (2008) Kenya: Floods, Appeal extension-recovery phase. Operations update 5. Available from: https://reliefweb.int/sites/reliefweb.int/files/resources/5B3384A4ACF23A70492574EF00257E51-Full_Report.pdf

Jahani A., Feghhi J., Makhdoum M. F., Omid M. (2016) Optimized forest degradation model (OFDM): An environmental decision support system for environmental impact assessment using an artificial neural networks. Journal of Environmental Planning and Management 59 (2), pp. 222 – 244.

Jahani A., Goshtasb H., Saffariha M. (2020) Tourism impact assessment modeling of vegetation density of protected areas using data mining techniques. Land Degradation & Development 31 (12), pp. 1502 – 1519.

Jamali S., Jönsson P., Eklundh L., Ardö J., Seaquist J. (2015) Detecting changes in vegetation trends using tine series segmentation. Remote Sensing of Environment 156, pp. 182 – 195.

Jamieson K., Talwalkar A. (2016) Non-stochastic best arm identification and hyperparameter optimization. Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS), Journal of Machine Learning Research 14.

Jebb A. T., Tay L., Wang W. Huang Q. (2015) Time series analysis for psychological research: Examining and forecasting change. Frontiers in Psychology 6: 727, pp. 1 – 24.

Jones H. G., Vaughan R. A. (2010) Remote sensing of vegetation: principles, techniques and applications. Oxford University Press, New York, p. 353.

Julien Y. (2008) Vegetation monitoring through retrieval of NDVI and LST time series from historical databases. PhD Dissertation, The University of Valencia, p. 320.

Justice C. O., Holben B. N., Gwynne M. D. (1986) Monitoring east Africa vegetation using AVHRR data. International Journal of Remote Sensing 7 (11), pp. 1453 – 1474.

Kalekar P. S. (2004) Time series forecasting using Holt-Winters Exponential smoothing. Report. Kanwal Rekhi School of Information Technology.

Karlik B., Olgac A. V. (2010) Performance analysis of various activation functions in generalized MLP architectures of neural networks. International Journal of Artificial Intelligence and Expert Systems 1 (4), pp. 111 – 122.

Karnieli A., Agam N., Pinker R. T., Anderson M., Imhoff M. L., Gutman G. G., Panov N., Goldberg A. (2010) Use of NDVI and land surface temperature for drought assessment: Merits and limitations. Journal of Climate 23 (3), pp. 618 – 633.

Key T., Warner T. A., McGraw J. B., Fajvan M. A. (2001) A comparison of multispectral and multitemporal information in high spatial resolution imagery for classification of individual tree species in a temperate hardwood forest. Remote Sensing of Environment 75 (1), pp. 100 – 112.

Kriegler F. J., Malila W. A., Nalepka R. F., Richardson W. (1969) Preprocessing transformations and their effects on multispectral recognition. Remote Sensing of Environment, VI, 97. Available from: https://ui.adsabs.harvard.edu/abs/1969rse..conf...97K

Kingma D. P., Ba J. L. (2015) ADAM: A method for stochastic optimization. Proceedings of the International Conference on Learning Representations.

Kinyanjui M. J. (2010) NDVI-based vegetation monitoring in Mau forest complex, Kenya. African Journal of Ecology 49, pp. 163-174.

Kipchirchir K. O. (2014) Comparative evaluation of six Indigenous Rangeland Grasses for Pasture Production under Varying Soil Moisture Contents in Tana River County, Southeastern Kenya. PhD thesis, University of Nairobi, Kenya.

Knoop L., Sambalino F., Van Steenbergen F. (2012) Securing water and land in the Tana Basin: A resource book for water managers and practitioners. Wageningen, The Netherlands: 3R Water Secretariat.

Kotu V. and Deshpande B. (2019) Data science: Concepts and practice. 2nd ed., San Mateo, CA, USA: Morgan Kaufmann (Publishers, Inc.).

Le Houérou H. N. (2009) Bioclimatology and biogeography of Africa. Heidelberg, Springer

Liashchynskyi P. & Liashchynskyi P. (2019) Grid search, random search, genetic algorithm: A big comparison for NAS. ArXiv, abs/1912.06059.

Lott F. C., Christidis N., Stott P. A. (2013) Can the 2011 east African drought be attributed to human-induced climate change? Geophysical Research Letters 40 (6), pp. 1177 – 1181.

Love E. R. (2015) Time series analysis of MODIS NDVI data with cloudy pixels: Frequency-domain and SiZer analyses of vegetation change in Western Rwanda. Master's Thesis, University of Tennessee, p. 76.

Malonza P. K., Malombe I., Muasya A. M., Lange C. N., Webala P., Mulwa R., Wasonga V. D., Mwachala G., Muasya J., Kirika P. M., MalakiP. (2006) Biodiversity assessment in dryland hilltops of Kitui and Mwingi Districts, Technical report, National Museums of Kenya, Nairobi, Kenya.

Mas J. F., Flores J. J. (2008) The application of artificial neural networks to the analysis of remotely sensed data. International Journal of Remote Sensing 29 (3), pp. 617 – 663.

McCleary R., Hay R., Meidinger E. E., MacDowall D. (1980) Applied time series analysis for the social sciences. California, SAGE Publications, p. 331

McKinney, W. (2010). Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference (SCIPY 2010), pp. 51 – 56. Available from: https://pandas.pydata.org/

Mehrotra K., Mohan C. K., Ranka S. (1997) Elements of Artificial neural networks. The MIT Press, Cambridge, MA.

Mayuge District Local Government (2020) March to May 2020 seasonal rainfall outlook over Uganda. Available from: https://mayuge.go.ug/march-may-2020-seasonal-rainfall-outlook-over-uganda

Mishra A. K., Desai V. R. (2006) Drought forecasting using feed-forward recursive neural network. Ecological Modelling 198, pp. 127 – 138.

Mishra A. K., Desai V. R., Singh V. P., F. ASCE. (2007) Drought forecasting using a hybrid stochastic and neural network model. Journal of Hydrologic Engineering 12 (6), pp. 626 – 638.

Muchayan A. (2019) Comparison of Holt and Brown's double exponential smoothing methods in the forecast of moving price for mutual funds. Journal of Applied Science, Engineering, Technology and Education 1 (2).

Mutti P. R., Lúcio P. S., Dubreuil V., Bezerra B. G. (2020) NDVI time series stochastic models for the forecast of vegetation dynamics over desertification hotspots, International Journal of Remote Sensing 41(7), pp. 2759 – 2788.

Musau, J., Patil S., Sheffield J., Marshall M. (2018) Vegetation dynamics and responses to climate anomalies in east Africa. Earth System Dynamics Discussions [preprint]

Nabavi-Pelesaraei A., Rafiee S., Hosseini-Fashami F., Chau K.-W. (2021) Chapter 11 – Artificial neural networks and adaptive neuro-fuzzy inference system in energy modelling of agricultural products. In Deo R., Samui P., Roy S. S. (eds), Predictive modeling for energy management and power systems engineering. Elsevier, pp. 299 – 334.

Nair V., Hinton G. E. (2010) Rectified linear units improve restricted Boltzmann machines. Proceedings of the 27th International Conference on Machine Learning.

NASA, Earth Observatory. (2011) Severe drought causes famine in east Africa. Available from: https://earthobservatory.nasa.gov/images/51411/severe-drought-causes-famine-in-east-africa

Nau R. (2020) Statistical forecasting: notes on regression and time series analysis. [Online Lecture notes]. Duke University. Available from: https://people.duke.edu/~rnau/411home.htm

National Drought Management Authority (NDMA). (2019) National drought early warning bulletin. Available from: https://www.ndma.go.ke/

National Oceanic and Atmospheric Administration (NOAA). (2021) What is remote sensing? Available from: https://oceanservice.noaa.gov/facts/remotesensing.html#:~:text=Remote%20sensing%20is%20the%20science,typically%20from%20aircraft%20or%20satellites.&text=Remote%20sensors%20collect%20data%20by,satellites%20or%20mounted%20on%20aircraft

NCSS Statistical Software (2021) The Box-Jenkins method. (Online notes). Available from: https://www.ncss.com/software/ncss/time-series-and-forecasting-in-ncss/

Ndungu L., Oware M., Omondi S., Wahome A., Mugo R., Adams E. (2019) Application of MODIS NDVI for monitoring Kenyan rangelands through a web based decision support tool. Frontiers in Environmental Science 7, 187.

Neal R. M. (1992) Connectionist learning of belief networks. Artificial Intelligence, 56 (1), pp. 71 – 113

Neupane A., Raj N., Deo R., Ali M. (2021) Chapter 6 – Development of data-driven models for wind speed forecasting in Australia. In Deo R., Samui P., Roy S. S. (eds), Predictive modeling for energy management and power systems engineering. Elsevier, pp. 143 – 190.

Nicholson S. E. (1996) A review of climate dynamics and climate variability in Eastern Africa. In: Johnson T. & Odada E. (eds). The limnology, climatology and paleoclimatology of the east African lakes. CRC Press, Boca Raton, pp. 25 – 56.

NIST/SEMATECH (2012) e-Handbook of methods. Available from: https://doi.org/10.18434/M32189

Nwankpa C. E., Ijomah W., Gachagan A., Marshall S. (2018) Activation functions: Comparison of trends in practice and research for deep learning. ArXiv, abs/1811.03378.

Obua J., Agea J. G., Ogwal J. J. (2010) Status of forests in Uganda, African Journal of Ecology 48 (4), pp. 853 – 859.

OECD/FAO (2016) Agriculture in sub-Saharan Africa: Prospects and challenges. OECD-FAO Agricultural Outlook, 2016 – 2025, 181. Available from: https://doi.org/10.1787/agr_outlook-2016-en

Oracle. (2020) Virtual machines. Oracle Cloud Infrastructure. Available from: https://www.oracle.com/cloud/compute/virtual-machines.html

Ornella L., Kruseman G., Crossa J. (2020) Chapter 1 Satellite data and supervised learning to prevent impact of drought on crop production: Meteorological drought in Gondrasek G. (ed), Drought, IntechOpen.

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E. (2011). Scikit-learn: Machine learning in python. Journal of Machine Learning Research 12, pp. 2825 – 2830.

Pellant M., Shaver P. L., Pyke D. A., Herrick J. E., Lepak N., Riegel G., Kachergis E., Newingham B. A., Toledo D., Busby F. E. (2020) Interpreting indicators of rangeland health. Version 5, Technical Reference 1734 – 6. U. S. Department of the Interior, Bureau of Land Management, National Operations Venter, Denver, CO.

Peng F., Quangang Y., Xue X., Guo J., Wang T. (2015) Effects of rodent-induced land degradation on ecosystem carbon fluxes in an alpine meadow in the Qinghai-Tibet Plateau, China. Solid Earth, 6, pp. 303 – 310.

Pettorelli N., Ryan S., Mueller T., Bunnefeld N., Jędrzejewska B., Lima M., Kausrud K. (2011) The normalized difference vegetation index (NDVI): Unforeseen successes in animal ecology. Climate Research 46, pp. 15 – 27.

Pettorelli N., Vik J. O., Mysterud A., Gaillard J-M., Tucker C. J., Stenseth N. C. (2006) Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends in Ecology and Evolution, 20 (9), pp. 203 – 510.

Pfeifer M., Burgess N. D., Swetnam R. D., Platts P.J., Willcock S. Marchant R. (2012) Protected areas: Mixed success in conserving east Africa's evergreen forests. PLoS ONE 7: e39337.

Pfeifer M., Platts P., Burgess N., Swetnam R., Willcock S., Lewis S., Marchant R. (2013) Land use change and carbon fluxes in east Africa quantified using earth observation data and field measurement. Environmental Conservation 40 (3), pp. 241 – 252.

Priddy K. L., Keller P. E. (2005) Artificial neural networks, an introduction. SPIE – The International Society of Optical Engineering, Washington, USA.

Riedmiller M., Braun H. (1992) A direct adaptive method for faster backpropagation learning: The RPROP algorithm. Proceedings of ISCIS VII, Universitat, Citeseer.

Rodrigues A. S. (2014) Analysis of vegetation dynamics using time-series vegetation index data from Earth Observation Satellites. PhD. Dissertation, Universidade do Porto, p. 156.

Rouse J. W., Haas R. H., Schell J.A., Deering D. W. (1974) Monitoring vegetation systems in the Great Plains with ERTS. Proceedings 3rd Earth Resources Technology Satellite-1 Symposium, NASA, SP – 351, Washington DC, pp. 309 – 317.

Roy B., (2021) A machine learning approach to monitoring and forecasting spatio-temporal dynamics of land cover in Cox's Bazar District, Bangladesh from 2001 to 2019. Environmental Challenges 5, 100237.

Rukikaire K. (2021) New atlas reveals rangelands cover half the world's land surface, yet often ignored despite threats. Available from: https://www.unep.org/news-and-stories/press-release/new-atlas-reveals-rangelands-cover-half-worlds-land-surface-yet

Rumelhart D. E., (1986) Parallel distributed processing, Explorations in the microstructure of cognition, volume 1: Foundations. The MIT Press, Cambridge, MA.

Running S. W., Justice C. O., Salomonson V., Hall D., Barker J., Kaufmann Y. J., Strahler A. H., Huete A. R., Muller J. -P., Vanderbilt V., Wan Z. M., Teillet P., Carneggie D. (1994) Terrestrial remote sensing science and algorithms planned for EOS/MODIS. International Journal of Remote Sensing 15 (17), pp. 3587 – 3620.

Ryan S. J., Cross P. C., Winnie J., Hay C., Bowers J., Getz W. M. (2012) The utility of normalized difference vegetation index for predicting African buffalo forage quality. The Journal of Wildlife Management 76 (7), pp. 1499 – 1508.

Saffariha, M., Jahani, A., Jahani, R., Latif S. (2021). Prediction of hypericin content in *Hypericum perforatum* L. in different ecological habitat using artificial neural networks. Plant Methods, 17, 10.

Sakamoto T., Yokozawa M., Toritani H., Shibayama M., Ishitsuka N., Ohno H. (2005) A crop phenology detection method using time-series MODIS data. Remote Sensing of Environment 96 (3 – 4), pp. 366 – 374.

Schoen-Phelan B. (2020) Chapter 3 – AstroGeoInformatics: From data acquisition to further applications. In Škoda P., Adam F. (eds), Knowledge Discovery in Big Data from Astronomy and Earth Observation. Elsevier, pp. 31 – 38.

Sharma N., Deo R. (2021) Wind speed forecasting in Nepal using self-organizing map-based online sequential extreme learning machine. In Deo R., Samui P., Roy S. S. (eds),

Predictive modeling for energy management and power systems engineering. Elsevier, pp. 437 – 484.

Shobel B. (2012). Stochastic projections of the financial experience of social security programs: Issues, limitations and alternatives. International Actuarial Association, Ontario, Canada.

Skipper S., Perktold J. (2010) Statsmodels: Econometric and statistical modelling with python, Proceedings of the 9th Python in Science Conference (SCIPY 2010). Available from: https://www.statsmodels.org/stable/index.html

Sruthi S., Aslam M. A. M. (2015) Agricultural drought analysis using the NDVI and land surface temperature data; A case study of Raichur district. Aquatic Procedia 4, pp. 1258 – 1264.

Stige L. C., Stave J., Chan, K-S., Ciannelli L., Pettorelli N., Glantz M., Herren H. R., Stenseth N. C. (2006) The effect of climate variation on agro-pastoral production in Africa. Proceedings of the National Academy of Sciences of the United States of America 103 (9), pp. 3049 – 3053.

Sun W., Song X., Mu X., Gao P., Wang F., Zhao G. (2015) Spatiotemporal vegetation cover variations associated with climate change and ecological restorations in the Loess Plateau. Agricultural and Forest Meteorology 209 – 210, pp. 87 – 99.

The European Space Agency. (2021). Earth from space: Tana River. Available from: https://www.esa.int/Applications/Observing_the_Earth/Earth_from_Space_Tana_River

Townshend J. R. G., Justice C. O. (1986) Analysis of the dynamics of African vegetation using the normalized difference vegetation index. International Journal of Remote Sensing 7(11), pp. 1435 – 1445.

Trodd N. M., Dougill A. J. (1998) Monitoring vegetation dynamics in semi-arid African rangelands: Use and limitations of earth observation data to characterize vegetation structure. Applied Geography 18, pp. 315 – 330.

Tucker C. J. (1977) Asymptotic nature of grass canopy spectral reflectance. Applied Optics 16 (5), pp. 1151 – 1156.

United Nations (2015) Transforming our world: The 2030 agenda for sustainable development. Available from: https://sdgs.un.org/goals

van Leeuwen W. J. D., Orr B. J., Marsh S. E., Herrnmann S. M. (2006) Multi-sensor NDVI data continuity: Uncertainties and implications for vegetation monitoring applications. Remote Sensing of Environment 100 (1), pp. 67 – 81.

van Rijn, J. N., Hutter F. (2018) Hyperparameter importance across datasets. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2369 – 2376.

Van Rossum G., Drake F. L. (2009) Python 3 reference manual, CreateSpace, Scotts Valley, CA. Available from: https://www.python.org/

Verdcourt B., Lind E. M., Morrison M. E. S., Hamilton A. C. (1975) East African vegetation, Kew Bulletin, 30 (1).

Wakabi W. (2006) Worst drought in a decade leaves Kenya crippled. Lancet, 367 (9514), pp. 891 – 892.

Walling D. E., Fang D. (2003) Recent trends in the suspended sediment loads of the world's rivers. Global and Planetary Change, 39 (1 – 2), pp. 111 – 126.

Wang S., Yang B., Yang Q., Lu L., Wang X., Peng Y. (2016) Temporal trends and spatial variability of vegetation phenology over the Northern Hemisphere during 1982-2012. PLoS ONE 11 (6): e0157134

Warmerdam F., Rouault E., et. al. (2021) GDAL. Available from: https://gdal.org/

Waswa B. S. (2012) Assessment of land degradation patterns in Western Kenya: Implications for restoration and rehabilitation. PhD Dissertation. ZEF. University of Bonn, Germany.

Weerts, H. J. P, Müller A. C., Vanschoren J. (2020) Importance of tuning hyperparameters of machine learning algorithms. ArXiv, abs/2007.07588

Wetlands International Kenya. (2015) The economics of ecosystem services of the Tana River Basin, mainstreaming the value of ecosystems and biodiversity in development planning. Available from: https://africa.wetlands.org/wp-content/uploads/sites/4/dlm_uploads/2017/08/TEEB-Study-Brief-Final-factsheet.pdf

Wichern J., van Wijk M. T., Descheemaeker K., Frelat R., van Asten P. J. A., Giller K. E. (2017) Food availability and livelihood strategies among rural households across Uganda. Food Security, 9, pp. 1385 – 1403.

Winters P. R. (1960) Forecasting sales by exponentially weighted moving averages. Management Science 6 (3), pp. 324 – 342.

Yao R., Wang L., Huang X., Chen X., Liu Z. (2019) Increased spatial heterogeneity in vegetation greenness due to vegetation greening in mainland China. Ecological Indicators 99, pp. 240 – 250.

Yengoh G. T., Dent D., Olsson L., Tengberg A. E., Tucker C. J. (2015) Applications of NDVI for land degradation assessment. Springer Briefs in Environmental Science, Springer, Cham.