MODELLING THE SPORADIC BEHAVIOUR OF RAINFALL IN THE LIMPOPO PROVINCE, SOUTH AFRICA: AN APPLICATION OF ETS STATE SPACE AND SARIMA MODELS

MASTER OF SCIENCE

in

STATISTICS

SV MOLAUTSI

2021

MODELLING THE SPORADIC BEHAVIOUR OF RAINFALL IN THE LIMPOPO PROVINCE, SOUTH AFRICA: AN APPLICATION OF ETS STATE SPACE AND SARIMA MODELS

by

SELOKELA VICTORIA MOLAUTSI

DISSERTATION

Submitted in fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

STATISTICS

in the

FACULTY OF SCIENCE AND AGRICULTURE
(School of Mathematical and Computer Sciences)

at the

UNIVERSITY OF LIMPOPO

SUPERVISOR: Prof. M Lesaoana

CO-SUPERVISOR: Dr. ME Moeletsi (Agricultural Research Council)

CO-SUPERVISOR: Dr. A Boateng (University of Cape Coast, Ghana)

2021

Declaration

I, **Selokela Victoria Molautsi**, declare that the research dissertation hereby submitted to the University of Limpopo, for the degree of Master of Science in Statistics has not previously been submitted for a degree at this or any other university; that it is my own work, and that all material contained herein has been duly acknowledged.

Molautsi, SV (Ms)	21/09/2021
.....	.....
Surname, Initials (title)	Date

Abstract

The effects of ozone depletion on climate change has, in recent years, become a reality, impacting on changes in rainfall patterns and severity of extreme floods or extreme droughts. The majority of people across the entire African continent live in semi-arid and drought-prone areas. Extreme droughts are prevalent in Somalia and eastern Africa, while life-threatening floods are common in Mozambique and some parts of other SADC (Southern African Development Community) countries. Research has cautioned that climate change in South Africa might lead to increased temperatures and reduced amounts of rainfall, thereby altering their timing and putting more pressure on the country's scarce water resources, with implications for agriculture, employment and food security. The average annual rainfall for South Africa is about 464mm, falling far below the average annual global rainfall of 860mm.

The Limpopo Province, which is one of the nine provinces in South Africa, and of interest to this study, is predominantly agrarian, basically relying on availability of water, with rainfall being the major source for water supply. It is, therefore, pertinent that the rainfall pattern in the province be monitored effectively to ascertain the rainy period for farming activities and other uses. Modelling and forecasting rainfall have been studied for a long time worldwide. However, from time to time, researchers are always looking for new models that can predict rainfall more accurately in the midst of climate change and capture the underlying dynamics such as seasonality and the trend, attributed to rainfall.

This study employed Exponetial Smoothing (ETS) State Space and Seasonal Autoregressive Integrated Moving Average (SARIMA) models and compared their forecasting ability using root mean square error (RMSE). Both models were used to capture the sporadic behaviour of rainfall. These two models have been widely applied to climatic data by many scholars and adjudged to perform creditably well. In an attempt to find a suitable prediction model for monthly rainfall patterns in Limpopo Province, data ranging from January 1900 to December 2015, for seven weather stations: Macuville Agriculture, Mara Agriculture, Marnits, Groendraal, Letaba, Pietersburg Hospital and Nebo, were analysed. The results showed that the two models were adequate in predicting rainfall patterns for the different stations in the Limpopo Province.

Dedication

This dissertation is dedicated to my lovely parents, grandfather, uncle and siblings for the outstanding support, prayers, encouragement and the love, who also taught me the value of education and critical thinking.

Acknowledgments

To start with, I would like to express my gratitude to my God in Heaven.
The timely and successful completion of the research could hardly be possible without the help and support from a lot of individuals. I take this opportunity to thank all those who assisted me directly or indirectly during this important work.

Firstly, I wish to express my sincere gratitude and due respect to my supervisor Prof Lesaoana. I am immensely grateful to her, for her valuable guidance, continuous encouragement and positive support which helped me a lot during the period of my work. I equally appreciate support offered by my co-supervisors: Dr Boateng and Dr Moeletsi. They always responded to my queries and provided valuable suggestions. I will forever be grateful for their contributions.
Secondly, I wish to express my whole-hearted thanks to my friends and classmates for their care and moral support. The moments I enjoyed with them during presentations and trips we had, will always remain as a happy memory throughout my life.

Thirdly, I owe a lot to my uncle for his constant love and support. He always encouraged me to have positive and independent thinking, which really matter in my life. I would like to thank him very much and share this joyful moment with him. Fourthly, I am grateful to Prof Nyamugure, a post-doctoral fellow who contributed towards my final editing of the dissertation, and the entire

staff members in the Department of Statistics and Operations Research for their unselfish and devotion towards me, over the course of my studies.

Last, but not least, the financial support received from the National Research Foundation through the University Capacity Development Grant of the National Graduate Academy for Mathematical and Statistical Sciences, Centre of Excellence in Mathematical and Statistical Sciences, and University of Limpopo Research Admin and Development, assisted in enhancing the quality of my research work. I greatly appreciate this assistance.

Contents

List of Figures

List of Tables

List of Abbreviations

ACF	Autocorrelation Function
ADF	Augmented Dickey-Fuller
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BIC	Bayesian Information Criterion
CCA	Canonical Correlation Analysis
EDA	Exploratory Data Analysis
ETS	Exponential Smoothing
GARCH	Generalised Autoregressive Condition Heteroskedasticity
GCMs	Global Climate Models
HEGY	Hylleberg Engle Granger Yoo
KPSS	Kwiatkowski Phillips Schmidt Shin
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error

MK	Mann Kendall
MLE	Maximum Likelihood Estimator
MSE	Mean Square Error
NRB	Nile River Basin
PACF	Partial Auto Correlation Function
PAR	Periodic Autoregressive
PP	Phillips Perron
RMSD	Root Mean Square Deviation
RMSE	Root Mean Square Error
SADC	Southern African Development Community
SARIMA	Seasonal Autoregressive Integrated Moving Average
SQMK	Sequential Mann Kendall
TSA	Theil Sen Approach
USAID	United States Agency for International Development
VAR	Vector Autoregressive
WN	White Noise

Chapter 1

Introduction and Background

1.1 Introduction

This chapter presents the background, problem statement, rationale, and the aim and objectives of the study. It also includes the methodology adopted and the scientific contributions of the study.

1.1.1 Background

The majority of people across the entire African continent live in semi-arid and drought-prone areas (Nicholson et al., 2018). Extreme droughts are prevalent in Somalia and eastern Africa, while life-threatening floods frequent Mozambique and some parts of other SADC (Southern African Development Community) countries.

Previous research already warned of increases in mean temperature over the years, resulting from climate change. Benhin (2006) indicated that climate

change in South Africa tends to increase temperatures, reduce the amount of rainfall, alter their timing and also put more pressure on the country's scarce water resources, with implications for agriculture, employment and food security.

The Department of Environmental Affairs (DEA) reported that during the period 1931-2015, South Africa warmed rapidly at the rate that exceeded 2^oC, mainly due to increased temperatures and water variabilities, with some parts of the country experiencing twice the global warming rate (Department of Environmental Affairs, 2018). The DEA further reported on evidence of increased extreme weather events in South Africa, which are observed through prolonged heat waves, dry spells and intensive rainfall (Department of Environmental Affairs, 2018). Extreme weather events include severe floods, storms and droughts, with drought experienced in recent years being declared as a national disaster (Department of Environmental Affairs, 2018). Over the next few decades, it is projected that South Africa will experience 4^oC to 6^oC rate of warming (Department of Environmental Affairs, 2018, 2019), hence a need to continuously monitor the changing effects in order to come up with measures that will mitigate the situation.

The African continent at large is prone to increased temperatures caused by global warming. The sub-Saharan region will be the hardest hit since it is most vulnerable to drought and climate conditions. In South Africa, the winter rainfall season is expected to be much drier than other countries in the region, and relatively high temperatures are already experienced, particularly in the Western Cape, Northern Cape, Limpopo and Mpumalanga. Again, Limpopo, Western Cape, Northern Cape, Gauteng and KwaZulu-Natal experience more than twice the global warming rate. Increases in extreme daily rainfall events are prevalent in Limpopo, North West, Free State, Gauteng and KwaZulu-Natal

provinces (Department of Environmental Affairs, 2018, 2019). This study has, therefore, targeted one of the most vulnerable regions to climate change, i.e. the Limpopo Province.

Not only South Africa, but also the sub-Saharan region will be affected, given that more than half of the region's staple maize is produced in South Africa. The most important factor limiting production from agriculture and other sectors in South Africa, is the availability of water. Rainfall is distributed unevenly across the country, with humid, subtropical conditions in the east and dry, desert conditions in the west (Benhin, 2006).

Statistically, it has been shown that South Africa has been getting hotter over the past four decades. The average rainfall in the country is very low, estimated at 450mm per year, well below the world's average of 860mm per year, and evaporation is comparatively high (Kruger and Shongwe, 2004). The surface and underground water are very limited, with more than 50% of the available water resources being used for only 10% of the country's agricultural activities (Benhin, 2006).

Characteristics and amount of rainfall are not easily known until rainfall actually occurs. Rainfall plays a crucial role in the evaluation and management of drought, floods and other related events, therefore, it is crucial to forecast rainfall (El-Shafie et al., 2012). Rainfall is natural climatic phenomena whose prediction is challenging and demanding (Nirmala and Sundaram, 2010). Weather parameters such as Rainfall, Relative Humidity, Wind Speed and Air Temperature are highly non-linear and complex, and require, among others, mathematical simulation and modelling for their accurate forecasts (Jain and Mallick, 2017). Time series models are appropriate tools for forecasting monthly rainfall, hence adopting these models in our study, i.e. to forecast the sporadic

behaviour of rainfall in the Limpopo Province. Changing climate conditions require scholars to pay more attention towards seeking optimisation precipitation methods that would ultimately achieve more accurate forecasts for agricultural and meteorology analysis (Valipour, 2016). More accurate forecasting of monthly rainfall is significantly important in water resource management and crop pattern design.

In this study, two-time series methods are employed to analyse the rainfall in the Limpopo Province: Seasonal Autoregressive Integrated Moving Average (SARIMA) and the Exponential smoothing (ETS) State Space models. ARIMA, also called the Box-Jenkins method, is a widely used time series model for analysing chronological data in different fields such as business statistics, meteorology, prediction technology, engineering and many more. Rainfall, the world over, is regarded as a seasonal phenomenon with the period of 12 months, and seasonal time series are modelled by Seasonal ARIMA technique which is a linear model; and ETS is a non-linear model (Etuk and Mohamed, 2014).

1.2 Problem Statement

Forecasting has become increasingly significant in formulating crucial estimates of things to come (Bako, 2014). The importance of time series in modelling and forecasting rainfall has been emphasised by many researchers (Papacharalampous et al., 2018; Crevits and Croux, 2017; Asemota, 2016; Marera, 2016; Murat et al., 2016; Ibrahim and Dauda, 2016). Hitherto strategists, policymakers, project managers and investors rely on forecasting for assistance in strategic planning, investment, policy planning, and resource mobilisation, in the short run. However, policymakers and planners are mindful that the basic and crucial purpose of forecasting is to predict in the near term, what will happen in order to avoid substantial cost or loss (Bako, 2014). Modelling and

forecasting the sporadic nature of rainfall has been carried out in a long time over the globe by several scholars (Subbaiah Naidu, 2016; Nicholls, 1980). The problem is to find an appropriate model that can capture both the dynamics such as the seasonality and the trend, attributed to rainfall.

From time to time, researchers develop new models that can predict rainfall patterns accurately. Nevertheless, South Africa, in particular, the Limpopo province being a water scarcity province with predominantly agricultural activities (Maponya and Mpandeli, 2012), has received less attention in terms of forecasting rainfall patterns. The socio-economic and agricultural activities in the Limpopo province of South Africa are mainly reliant on rainfall. Accordingly, the production from the agricultural sector each year is largely determined by the rainfall seasonal pattern, trend and cycles, hence modelling the sporadic or seasonal behaviour of rainfall has become more imperative today than ever before, especially in the wake of climatic change. Owing to the agricultural base in the Limpopo province, it is important to model seasonal behaviour of rainfall in the province for agricultural planning. This study will examine the sporadic behaviour of monthly rainfall for the following stations of the Limpopo province: Macuville Agriculture, Mara Agriculture, Marnits, Groendraal, Letaba, Pietersburg Hospital and Nebo, using the error trend seasonal (ETS) state space and seasonal autoregressive integrated moving average (SARIMA) models.

1.3 Rationale

Observations on rainfall in South Africa already show great inter-annual variabilities that are prone to cause either drought- or flood-related weather events (Schulze, 2003). The South African Weather Service noted that although South

Africa's average rainfall has been low, only certain areas of the country experience a meteorological drought.

Rainfall is the most important factor in both atmospheric and climatological cycles. Rainfall plays a major role in hydrology that finds its greatest applications in the design and operations of water resources, engineering works as well as agricultural systems. Effective planning therefore, requires cautious analysis of the behaviour of temporal rainfall (Yasmeen and Hameed, 2018). Subbaiah Naidu (2016) noted that the analysis of rainfall records for long periods provides information about rainfall patterns and variability.

Chonge et al. (2015) commented that time series analysis provides a great opportunity for describing, explaining, modelling and predicting climatic variability and its impacts. To understand meteorological information and integrate it into planning and decision-making process, it is crucial to study the temporal characteristics and predict lead times of rainfall in the Province. Rainfall forecast has become paramount in the wake of changing climatic conditions due to human activities (Ampaw et al., 2013). Even though devastations resulting from extreme rainfall cannot be avoided completely, a forewarning could be useful (Takele, 2012; Nicholls, 1980). Therefore, there is a need to study the impact of sporadic rainfall in detail.

The dynamic nature of weather resulting in climate change, has significantly affected the rainfall patterns worldwide (Ampaw et al., 2013). According to Nieuwolt et al. (1983), the non-static behaviour in rainfall patterns comes in various magnitudes ranging from variability through fluctuations, trends, and abrupt to gradual changes. In South Africa, climate variability and change are two key concerns as they pose a significant threat to water resources, food security, health, infrastructure, as well as ecosystem services and biodiversity

(Ziervogel et al., 2014).

According to Ampaw et al. (2013), the lifeblood of every living organism is water. Rainfall being one source, plays a critical role in various aspects of our physical, chemical and biological processes, and also serves as a regulator of the worldwide environment (Jury and Vaux, 2005). Some features of rainfall which include its seasonal or sporadic behaviour and day-time distribution, intensity, duration, onset, cessation and frequency of rain days, all show important variations in respect to time and places (Cook and Heerdegen, 2001). It is, therefore, essential to examine rainfall to determine possible trends and changes in the data generating process.

Takele (2012) asserts that rainfall typically has spatial and temporal variability, which may affect agricultural output, water supply, transportation, environment, urban planning, or the entire economy of a country. These natural phenomena may negatively affect the agricultural production and/or the economy at large (Takele, 2012). Considering South Africa's high levels of poverty and inequality, these impacts constitute critical challenges for national development (WIREs Climate Change, 2014).

Htike and Khalifa (2010), and Subbaiah Naidu (2016), among others, have noted that predicting the trend of rainfall is a difficult task in meteorology and environmental sciences. The problem at stake is, therefore, how to accurately predict seasonality and trend in rainfall in all the selected stations of the Limpopo province for farmers and other policy makers.

In South Africa, few studies have been conducted to analyse rainfall data using ETS state space model. Again, most research conducted on rainfall mainly relied on linear models such as ARIMA and SARIMA, amongst others (Sub-

baiah Naidu, 2016; Chonge et al., 2015). These models are often incapable of modelling non-linear behaviour, which are often related to rainfall data (Yusof and Kane, 2013). Therefore, this study will employ a non-linear model together with a linear model and compare their forecasting ability using MSE (Mean square error) and RMSE (Root mean square error), among others.

1.3.1 Research Questions

This study sought to answer the following questions:

1. What is the relationship between rainfall patterns in the selected stations of the Limpopo province?
2. Which model between ETS state space and SARIMA can best forecast the sporadic rainfall in the selected stations of the Limpopo province?

1.3.2 Aim

The aim of the study is to model sporadic rainfall behaviour in the selected stations of the Limpopo province, using the Exponential smoothing (ETS) state space and SARIMA models.

1.3.3 Objectives

The objectives of the study are to:

1. determine the relationship between rainfall pattern in selected stations of the Limpopo province.
2. propose a suitable model for forecasting the sporadic rainfall in all the selected stations of the Limpopo province using ETS state space and SARIMA models.

3. compare the forecast ability of ETS state space and SARIMA models.

1.4 Methodology and Analytical Procedures

1.4.1 Data Collection

Secondary data has been used for this study. A time series of daily and monthly rainfall data in millimetres (mm) for the period from January 1900 to December 2016 was conducted. The data was obtained from the Agricultural Research Council (ARC) for the following selected weather stations in the Limpopo province of South Africa: Macuville Agriculture, Mara Agriculture, Marnits, Groendraal, Letaba, Pietersburg Hospital and Nebo. Although the data was obtained from the ARC, it emanates from the South African Weather Service (SAWS). Details of how the SAWS collects the data are available on their website (www.weathersa.co.za).

1.4.2 Data Analysis

We performed an exploratory data analysis using a series of plots and graphs. We then established stationarity using standard tests such as Augmented Dickey-Fuller (ADF), Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and Phillips-Perron (PP) tests (Kwiatkowski et al., 1992; Perron, 1989; Dickey and Fuller, 1981, 1979). Stationarity in time series data is the first important step in modelling and forecasting. Rainfall data often exhibits sporadic or seasonal behaviour and this has been ascertained using the HEGY (Hylleberg-Engle-Granger-Yoo) test (Hylleberg et al., 1990). After checking all the aforementioned, SARIMA and ETS state space models were fitted to the data (Subbaiah Naidu, 2016; Hyndman et al., 2002). Indeed, the best amongst SARIMA and ETS state space models was selected using the Akaike Information Criteria (AIC) (Akaike, 1978). These two models were then used to forecast out-of-sample rainfall patterns

in all the selected stations of the Limpopo province. The forecasting ability between SARIMA and ETS state space models were compared using MSE, RMSE, among others. We used R and SPSS statistical software packages and both models were estimated using the maximum likelihood estimator.

1.5 Scientific Contribution

The outcomes of this research study were used for forecasting the pattern of rainfall in the selected stations of the Limpopo province. The study provides information such as early warning system in the province, that would be helpful for decision makers in formulating policies to mitigate the problems of rain water resources management, soil erosion, flooding and drought. Eventually, the forecasting model developed will be a valuable instrument for the agricultural, hydrological and socio-economical sectors. The results could also be used to evaluate rainfall extreme risk and contribute to the development of mitigation strategies.

1.6 Organisation of the Study

This study is divided into five chapters. Following this introductory chapter, is Chapter 2 that provides an overview of the related works. Chapter 3 discusses the statistical methods used in this study for modelling and forecasting rainfall. Chapter 4 presents the results and discussion of the study and Chapter 5 summarises the dissertation, draws the appropriate conclusions and recommendations; and outlines some potential directions for further research work.

1.7 Chapter Summary

The chapter gave an introduction to the study, highlighting on issues relating to background, problem statement, aim and the objectives of study, methodology and scientific contribution.

Chapter 2

Literature Review

2.1 Introduction

This chapter firstly looks at the general definitions of rainfall, followed by time series analysis. The chapter also provides selective references and definitions for a clearer understanding of the contributions of Box-Jenkins methodology and ETS state space exponential smoothing models in time series modelling and forecasting.

2.2 Definition and Concepts

2.2.1 General

Several scholars have studied changes brought by global warming that have impacted on the frequency and intensity of meteorological parameters such as extreme floods, droughts, temperature, heat waves, humidity and wind speed (Abdullah et al., 2019; Salhi et al., 2019; Caloiero et al., 2018; Lavanya et al.,

2018; Dosio, 2017; Kruger and Nxumalo, 2017; Takele and Gebretsidik, 2015). Weather and climate over the earth are not constant with time, but vary from decade to decade and change on different time series, ranging from the geological to diurnal through intra-seasonal, annual and seasonal scales (Takele, 2012).

While Kruger and Nxumalo (2017) also observed significant climatic changes in recent years across the globe, they have noted some consistency in temperature increases as opposed to precipitation changes that have exhibited more variabilities.

Climate change is a debatable subject these days. Dealing with this topic is considered an enormous challenge of the coming years (Abdullah et al., 2019). There are many definitions of climate; however, the common term of climate is the long-term pattern of meteorological conditions in a specific area (Swain et al., 2015). It is measured by evaluating variations in temperature, humidity, atmospheric pressure, wind, precipitation, atmospheric particle, and other meteorological variables. Climate change can have significant impacts on weather conditions around the world, such as storms and heavy rainfall.

Climate in Africa is already exhibiting significant changes as evidenced by the changes in the amount of rainfall, changes in average temperature, patterns, the prevalence of frequency and intensity of weather extremes (Kotir, 2010). Shongwe et al. (2009) noted that the variability of rainfall makes the Southern Africa region vulnerable to extreme climate events. Southern Africa is characterised by significant climate variability on a range of temporal and spatial scales which impact on the economy, agriculture and water resources at large (Phakula et al., 2016).

In the Limpopo Province climate change is becoming increasingly apparent and the usual manifestation of climate change are evidenced by the long term changes in weather indicators like temperature and rainfall (Limpopo Environmental Outlook Report, 2016). Limpopo Environmental Outlook reports on the Limpopo climate that is divided into four climatic regions, i.e., the subtropical Lowveld region of hot-rainy summers and warm-dry winters, also known as the South African Bushveld; the moderate eastern plateau with warm to hot and rainy summers and cold dry winters; the escarpment region with colder weather because of the altitude and rain all year around; and the subtropical plateau with a flat elevated interior area, hot and dry with winter rain. Rainfall results from a series of complex interactions that take place within the earth's atmosphere system. Water, in all its forms and various activities, plays a crucial role in sustaining both the climate and life (Takele, 2012).

2.2.2 Rainfall Characteristics

South Africa is situated at the latitude of 22°–34° south and longitude of 16°–32° east. The country is bordered by the Indian Ocean on the south, the Atlantic Ocean on the west and at the east with a coastline stretching over 2500km. On its northern border are Namibia, Botswana, and Zimbabwe, and to the north-east are Mozambique and Swaziland. Surrounded by South Africa and on its eastern central part is the independent Kingdom of Lesotho (Palmer and Ainslie, 2002). The country is divided into nine administrative provinces: North West, Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, Northern Cape, and Western Cape.

Geographically, South Africa is composed of three primary regions: the highveld comprised mainly of grassland; a nearly continuous escarpment of mountain ranges that encompasses the plateau and large central plateau with heights of 1220m to 1830m above sea level on the west, south and east, with heights

exceeding 3050m and its highest point in the Drakensberg mountains; and a narrow strip of low-lying land along the coast (South Africa, 2019).

The average rainfall for South Africa is about 450mm per year, which is well below the world average of about 860mm per year, and the rainfall is mostly received during austral summer seasons (November to March), except for the south western and south coast which receives its rainfall during austral winter and throughout the year (Philippon et al., 2011).

Malherbe et al. (2012) noted that there has been a perceptible decrease in the total rainfall on much of the eastern part of Southern Africa, including most of the Limpopo River Basin, and rainfall in the province varies significantly between the years. Typical rainfall for the Limpopo Province ranges from 200mm in the hot dry areas to 1500mm in the high rainfall areas, with most of it happening between October and April.

2.2.3 The Main Effects of Rainfall

The extreme rainfall (floods), soil erosion caused by floods in most cases and lack of rainfall (drought) are the events that cause significant damage to agriculture, ecology, infrastructure, and many others. These also cause injury, loss of lives and disruption to human activities. Socio-economic activities including water supply, power generation, human health and agriculture, among others, are sensitive to climate variability (Takele, 2012). Next, we elaborate on flood, drought and soil erosion.

Flood

Flood is an overflow of water onto normally dry land, the inundation of a normally dry area caused by rising water in an existing waterway such as a drainage ditch, stream or river, then ponding of water at or near the point

where the rain fell (National Weather Service, 2019). Flooding is a longer term event than flash flooding. These floods include river floods, flash floods, urban floods and coastal floods, among others (Kundzewics et al., 2014).

Osti et al. (2008) noted that the floods are the most frequent and devastating natural disaster in both the developing and developed countries. The torrential rains and flooding affected 600,000 people in 16 West African nations in September 2009 and the worst hit countries were Burkina Faso, Senegal, Ghana and Niger (Baldassarre et al., 2010). This event closely followed the 2007 floods that displaced more than a million people in Uganda, Ethiopia, Sudan, Burkina Faso, Togo, Mali and Niger, and claimed over 500 lives and the 2008 flooding in Mozambique (United Nations, 2009). Tropical Cyclone Dineo was one of the deadliest tropical cyclones on record in the South-West Indian Ocean and Southern Hemisphere as a whole (FloodList, 2017). It was the first tropical cyclone to hit Mozambique since Cyclone Jokwe in 2008. The following areas were affected by Cyclone Dineo: Mozambique, Limpopo and Mpumalanga (FloodList, 2017). Recently, tropical Idah made landfall during the night of 14 March 2019 near Beira City Mozambique, bringing torrential rains and strong winds, then moving westwards over east Zimbabwe and also causing massive flooding in Malawi and the loss of life and devastation in its trail (FloodList, 2019).

Southern Africa lies within the arid to semi-arid climatic regions and extreme rainfall events are frequent in the region (Lindesay and Jury, 1991; Mason and Jury, 1997). Alexander (1993) noted the possibility of collapses of small dams and increase in reservoirs, which could affect up to 100 000 people in South Africa settled on the floodplains which could be inundated, therefore the long-term implications of increases in the frequency and magnitude of high rainfall events pose serious problems.

In December 2010, South Africa experienced unusual weather patterns caused by the La Nina effect, resulting in floods that in turn led to unprecedented disruption of services, displacement of people, loss of livelihoods and even worse, loss of life (Disaster Relief Emergency Fund operations, 2011).

Figure 2.1: A flooded road in Gauteng Province-December 2010.

All the nine provinces of South Africa (Eastern Cape, Free State, Gauteng, Kwa-Zulu Natal, Limpopo, Mpumalanga, Northern Cape, North West and Western Cape) were affected by the floods (Disaster Relief Emergency Fund operations, 2011).

Figure 2.1 demonstrates the devastations created by the December 2010 floods in some part of the Gauteng province.

During 01-18 March 2014, the north-eastern part of South Africa experienced heavy and extended rainfall which led to extensive flooding and landslides, and affected over 7,000 people, with 3,525 displaced. The government of South Africa confirmed 32 deaths with others still missing due to flash flooding (Disaster Relief Emergency Fund operations, 2014). The most affected provinces were: Limpopo, Mpumalanga, North West and Gauteng.

Figure 2.2: Informal settlements washed away in Lephalale, Limpopo Province-March 2014.

Figure 2.3: Flooding in Bela-Bela, Limpopo Province-March 2014.

Limpopo Province was the worst affected and the following areas were hardest hit: municipalities of Lephalale, Mogalakwena, Modimolle, Bela-Bela, Thabazimbi, and Mookgopong in the Waterberg District. The cost of damage to infrastructure (main roads, houses, farmland and tourist centres) was estimated in millions of US dollars (Disaster Relief Emergency Fund operations, 2014).

Figure 2.2 and Figure 2.3 display the destructions caused by the March 2014 floods in some parts of the Limpopo province.

Short duration-intense rainfalls result in floods that can potentially cause se-

rious problems to urban infrastructure (i.e., dams, bridges, and storm water systems), population of informal settlements located in floodplains, and agriculture. The impacts and economical losses of extreme events have led to some scholars suggesting that climate change analysis be focused on changes in extreme events rather than changes on climatic means (South African Weather Service, 2015; Von Storch and Zwiers, 1988).

Drought

Drought is a period of time without substantial rainfall that persists from one year to the next, and it is a normal part of virtually all climatic regions, including areas with high and low average rainfall. Drought is the consequence of anticipated natural precipitation reduction over an extended period of time, usually a season or more in length (Concho Valley Council of Governments, 2012). Droughts can be classified as agricultural meteorological, hydrologic and socioeconomic.

The African continent is home to over one billion people, most of whom reside in semi-arid drought prone regions (Nicholson et al., 2018). The great concern is the projection that Africa may be the continent most impacted by climate change and that droughts (or lack of rainfall) could become more intense and frequent in some areas (Gizaw and Gan, 2017; Niang et al., 2014). Many researchers now believe that the occurrence of various droughts in Africa, especially in southern Africa and the Horn, are caused by physical processes related to the occurrence of ElNiño Southern Oscillation (ENSO) events (Wolde-Georgis, 1997). ENSO is a coupled air and ocean phenomenon with global weather implications (Bekele, 1997).

South Africa is prone to drought, and droughts in South Africa have affected the national economy and local communities (Baudoin et al., 2017). The drought

has created mounting pressures on the nation's agro-economic system, including increased unemployment, negative impacts on upstream economic activities (e.g., less purchase power) and increased debt service costs for farming enterprises (AgriSA, 2016). Citizens of neighbouring countries (Zimbabwe, Lesotho, Namibia, Botswana and Swaziland) that rely on food import from South Africa are also affected by reduced agricultural productivity in South Africa (United Nations Office, 2016). The major drought periods in recent history include 1982–1984, 1991–1992 and 1994–1995, and during these years, devastating environmental and socioeconomic impacts occurred in southern Africa (Unganai and Kogan, 1998). A severe drought experienced by South Africa in 2015–2016, seriously depleted the water reserves (Baudoin et al., 2017).

Otto et al. (2018) noted that from 2015 to 2017 (three consecutive years), the Western Cape province region received below average rainfall, leading to a prolonged drought and acute water shortages, most prominently in the City of Cape Town. Otto et al. (2018) further noted that the water crisis associated with the drought was very extreme at the beginning of 2018 to the point that the City of Cape Town was expected to run out of water by March 2018 (Figure 2.4 depicts the situation in some parts of Cape Town during the indicated period).

Figure 2.4: Dried out branches are seen amongst caked mud at Theewaterskloof dam near Cape Town, Western Cape-January 2018.

Limpopo is the breadbasket and agricultural engine of South Africa, accounting for nearly 60% of all fruit, vegetables, maize, wheat, and cotton. Livestock farming is also a significant contributor to the province's agriculture sector (Maponya and Mpandeli, 2016). The agricultural sector in South Africa, which uses more than 60% of fresh water, mostly towards irrigation, has been the hardest hit by the drought (Statistics South Africa, 2007). Extreme drought in the Limpopo River Basin is also a regular phenomenon and has been recorded for more than a century at intervals of 10-20 years, although the periodicity of droughts is not necessarily so predictable (FAO, 2004).

Soil Erosion

Soil erosion is a gradual process that occurs when the impact of water or wind detaches and removes soil particles, causing the soil to deteriorate. The soil erosion by water and the sediment attached nutrients, deteriorate water quality in the surface water body (Uddin et al., 2015). Figure 2.5 displays the eroded area caused by soil erosion in November 2014 at the Springbok flat, in the Limpopo Province.

Figure 2.5: Erosion on the Springbok flats, Limpopo Province-November 2014

Soil erosion is a major problem confronting the land and water resources throughout South Africa. Previous researches indicate that more than 70% of land in South Africa is affected by varying intensities of soil erosion (Le Roux, 2011; Le Roux et al., 2008).

Le Roux (2011) stated that soil erosion does not only include the loss of fertile topsoil and reduction of soil productivity, but it is also coupled with serious off-site impacts related to increased mobilisation of sediment and delivery to rivers. The eroded soil material leads to sedimentation/siltation of reservoirs, as well as an increase in pollution due to suspended sediment concentrations in streams, which in turn affects water use and ecosystem health (Flugel et al., 2003). This has resulted in a loss of productive soil from crop and grazing land, as well as layers of infertile soil being deposited on formerly fertile crop lands, the formation of gullies silting of lakes and streams, and land slips (Takele, 2012). The Eastern Cape, Limpopo and KwaZulu-Natal provinces have the highest erosion potential (Le Roux, 2011).

2.3 Time Series Analysis

The term time series denotes a data storage format which consists of two mandatory components: the time units and the corresponding values assigned for the given time unit. The values of the series must denote the same meaning and correlate among the nearby values (Ostashchuk, 2017). Making predictions or understanding the mechanisms that are generated at a time series data are among the main goals of time series analysis. The modelling of time series data is a subject of great importance in a variety of fields, e.g., business, agriculture, engineering, medicine, epidemiology, meteorology, and many others. Time series is in fact a stochastic procedure that describes the evolution of the random variable and its unit of time will mostly vary with the application. It might be in years, quarters, months, days, hours, minutes or even microseconds, depending on the situation to be modelled.

Iffat (2009) noted that the unit of time is not important, but what matters is that the observation be equally spaced in time. In time series studies the interest is on time delay or time lag, not the actual time. If the observations are not equally spaced, everything gets more complicated. The main aim of time series analysis is to develop a model that signifies time series, after which the model would be used to forecast the near future values. In general, the model would be used to predict future values of the given variable according to its past behaviour (Box and Jenkins, 1970).

2.4 Modelling and Forecasting Rainfall Time Series

Modelling and forecasting are some of the main aims of time series analysis. Time series analysis comprises of methods for analysing time series data in or-

der to extract meaningful statistics and other characteristics of the data. Time series data have a natural temporal ordering, which makes time series analysis different from other common data analysis problems, in which the ordering of the observations does not matter. Different approaches may be used for modelling and forecasting time series. Forecasts from the Box–Jenkins methodology are generally very good for short-term forecasting. Regression analysis, which can be used with moving average models or a series with deterministic trends, may also serve to make predictions over the longer run.

There are also the exponential smoothing method, the Kalman filter method and the Artificial neural network method, among others. Modelling and forecasting of rainfall dynamics have been undertaken for a long period of time, for example, Bagirov et al. (2017) presented a technique called Clusterwise Linear Regression method, which was used to predict monthly rainfall in Victoria, Australia using rainfall data with five input meteorological variables over the period 1889-2014 from eight weather stations. The proposed method was also compared with the multiple linear regression, artificial neural networks and the support vector machines for regression models, and the results demonstrated that the proposed method outperforms the other methods.

Zhou et al. (2014) evaluated the multifractal properties of daily rainfall time series at the stations in Pearl River basin of China over 45 years. They used the universal multifractal approach based on the multiplicative cascade model and the multifractal detrended fluctuation analysis. The results showed that the empirical K(q) curves of the daily rainfall time series can be fitted well by the universal multifractal approach.

Higareda et al. (2014) analysed rainfall variable trends and potential vegetation responses, using monthly data from 55 meteorological stations from the

North Pacific Watershed in Sinaloa, Mexico. Seven annual rainfall trends were estimated and analysed using non-parametric tests. Their study found that the indicated divergent trends in all the proposed variables, in general, increased and drier trends were found. Latitude was the most relevant factor regarding trend behaviour in geographic terms.

Sayemuzzaman and Jha (2014) used the non-parametric Mann-Kendall (MK) test, the Theil-Sen approach (TSA) and the Sequential Mann-Kendall (SQMK)) test to analyse the 249 stations of precipitation data in the state of North Carolina in the United States, over the period 1950-2009. Regional trend analysis found increasing precipitation in mountain and coastal regions. Piedmont region was found to have increasing trends in summer and fall, but decreasing trend in winter, spring and on an annual basis. The SQMK test on trend shift analysis, identified a significant shift during $1960-1970$ in most parts of the state. The comparison between winter (summer) precipitations with the North Atlantic Oscillation (Southern Oscillation) indices concluded that the variability and trend of precipitation can be explained by the Oscillation indices for North Carolina.

Bari et al. (2016) used the three non-parametric tests which were MK test, the TSA approach and the SQMK test to analyse the seasonal and annual rainfall trends in the northern region of Bangladesh, the period of the data set was 1964-2013 (50 years). The results showed that the SQMK test detected several non-significant points of change for seasonal and annual rainfall at most of the stations. Periodic fluctuations were also detected, and it was observed that for the majority of the stations in the region, there were decreasing seasonal rainfall trends after the early 1990s.

Longobardi and Villani (2010) used time series to analyse rainfall over a wide

time interval and a wide area detecting potential trends and assessing their significance in the Mediterranean area. They analysed the data set of 211 stations during the period 1918-1999. Their study established that the trend appeared predominantly negative, both at the annual and seasonal scales, except for the summer period when it appeared to be positive. Over the whole reference period, positive and negative trends were significant, respectively for 9% and 27% of total stations; and over the last 30 years a negative trend was instead significant for 97% of the total stations.

Onyutha et al. (2015) conducted a study to analyse rainfall trends in the Nile River Basin (NRB) in Africa. The non-parametric Mann-Kendall test was used to analyse the data from 39 locations of the NRB. In the equatorial region, in 4 of the 7 stations, trends in annual rainfall were found mainly positive and significant at level α = 5%. Selectively for Sudan, Ethiopia and Egypt, trends in the annual rainfall were mostly negative and significant at α = 5% in 69% of the 32 stations, and heterogeneity in the trend directions for the entire NRB was confirmed at α = 1% in 13% of the 39 stations.

Mazvimavi (2010) investigated the changes over time of annual rainfall in Zimbabwe. The scholar employed the MK test to analyse the data set for 40 rainfall stations for the period 1892–2000 (108 years). The results of the study showed that the long-term average input of rainfall into the land phase of the hydrological system had not changed and anthropogenic changes of land use and land cover accelerated soil erosion, resulting in increased siltation rates which, in some cases, adversely affected the available water resources.

Akinyemi et al. (2013) conducted a statistical analysis of annual and monthly rainfall patterns in Ekiti State, Nigeria. Descriptive and time series analyses were used to analyse the temporal distribution of rainfall for the period of

10 years (2001–2010). Their results showed a significant change in the distribution and characteristics of rainfall, such as occurrence and intensity in the monthly and annual rainfalls in Ekiti State. The results of the standardised anomaly Moving Average and the linear trend detected fluctuations in the annual rainfall, even though the positive and negative deviations were evenly distributed.

Jaiswal et al. (2015) conducted a statistical analysis for change detection and trend assessment in climatological parameters in Switzerland. The change detection analysis was conceded using Pettitt's test, Von Neumann ratio test, Buishand's range test and Standard Normal Homogeneity test; and non-parametric tests applied for the trend analysis, included linear regression, Mann-Kendall and Spearman rho tests. The results of the meteorological variables indicated different change points between the years 1990 to 2000, with maximum change points in and around 1995.

Stager et al. (2013) studied the late Holocene precipitation variability in the summer rainfall region of South Africa. The records confirmed that local tree ring and stalagmite gray scale series represent rainfall variability, but widely cited stable isotope series from Makapansgat do not represent past climate clearly. Many interpretations of the climatic history of southern Africa were influenced by the isotope data, re-examined late Holocene precipitation variability in the summer rainfall zone and the addressed model projections of future precipitation in the region.

Du Plessis and Burger (2015) investigated increasing short-duration rainfall intensities in South Africa. A sub-daily rainfall for three stations in the Western Cape and four stations in the rest of South Africa were analysed in order to determine if any trends towards more intense and extreme rainfall existed, and

whether the trend was unique to the Western Cape or indicated a wider trend. The magnitude of rainfall events, extreme value theory was applied to non-stationary sequences using both parametric and non-parametric approaches for event maxima and peaks over threshold modelling. The frequency analysis entailed measuring the frequency of exceedance of rainfall events over a certain threshold value. Both the magnitude and frequency analysis indicated that the combination of the two record types influenced the results of some of the stations, while the other stations showed no consistent evidence of changing rainfall intensities. The study then concluded that the observed short-duration record showed no evidence of trends or indications of changes in rainfall intensities.

Walker and Tsubo (2003) estimated rainfall intensity for potential crop production on clay soil with in-field water harvesting practices in semi-arid area of South Africa. The two successful models for rainfall-runoff processes were introduced to be used in crop model simulation: (1) the area under the rainfall intensity curve method and (2) the combination of dimensionless hyetograph model and deterministic runoff model. It was concluded that the latter is more useful than the former. This stochastic rainfall intensity model was developed to provide long-term rainfall intensity values as applied in water harvesting projects. The results were then available at the hydrologists, to be used in peak rainfall intensity estimation (for food production) and soil erosion and water quality studies.

Phakula et al. (2016) modelled the seasonal rainfall characteristics over South Africa. Global climate models (GCMs) are assessed over South Africa and used as forecasting skill in predicting seasonal characteristics. The GCMs output is configured to predict the number of rainfall days at South African Weather Service stations, exceeding pre-defined threshold values for the austral summer

seasons and to predict the rainfall totals of the onset months of the rainy seasons, for eight homogeneous rainfall regions of South Africa. Using canonical correlation analysis (CCA) as statistical downscaling technique through model output statistics, the forecast skill levels of coupled ocean–atmosphere and uncoupled atmospheric models were determined through retro-actively generated hindcasts. It was found that the CCA patterns for both the GCMs indicated that usually when there are anomalously negative (positive) predicted 850 hPa geopotential heights over South Africa, there are anomalously wet (dry) conditions over most parts of South Africa, and this work has paved the way for the operational production of seasonal rainfall characteristics over South Africa in real time.

2.5 Box-Jenkins Methodology on Rainfall Time Series

Box-Jenkins Analysis refers to a systematic method of identifying, fitting, checking, and using integrated autoregressive, moving average time series models. Autoregressive Integrated Moving Average Model (ARIMA) is a widely used time series analysis model for analysing chronological data in statistics. The ARIMA model was firstly proposed by Box and Jenkins in the early 1970s, hence often termed Box-Jenkins model or B-J model, for simplicity (Stoffer and Shumway, 2010). ARIMA is a type of short-term prediction model in time series analysis, because this method is relatively systematic, flexible and can grasp more original time series information (Cryer and Chan, 2008; Kantz and Schreiber, 2004).

The Box-Jenkins method refers to the iterative application of the following three steps: (1) Identification: Using plots of the data, autocorrelations, partial autocorrelations, and other information, a class of simple ARIMA models

is selected. This amounts to estimating appropriate values for p, d, and q. (2) Estimation: The ϕs and θs of the selected model are estimated using maximum likelihood techniques, backcasting, etc., as outlined in Box and Jenkins (1976). (3) Diagnostic checking: The fitted model is checked for inadequacies by considering the autocorrelations of the residual series (the series of residual, or error, values). These steps are applied iteratively until step (3) does not produce any improvement in the model. We then can forecast the data.

The periodicity of the time series is usually due to seasonal changes like quarterly, monthly and degree of weeks or some other natural changes. If a time series has seasonal variation, there will be autocorrelation at lag S and possibly at multiples of that lag, depending on the persistence and nature of that autocorrelation. This seasonal dependence can be described by Seasonal ARIMA (SARIMA) model of order (P,D,Q) if it is necessary to take seasonal difference (Subbaiah Naidu, 2016).

The ARIMA model is the most widely used method in meteorology, engineering technology, Marine, business statistics and prediction technology, for example, Weesakul and Lowanichchai (2005) developed ARIMA models that are most appropriate to forecast annual rainfall in all regions of Thailand with acceptable accuracy, and which are able to fulfil the requirement for agricultural water allocation planning.

Wang et al. (2013) applied the seasonal time series model in the precipitation forecast of Shouguang city, in China. The SARIMA model was used to model and forecast the data. It was found that rainfall has a strong autocorrelation of seasonal characteristics in time series, utilising seasonal periodicity with a SARIMA methodology. The researchers analysed the statistical data of precipitation which was based on Shouguang city in Shandong statistical

yearbooks over the 1996–2009 periods. Basically, it was concluded that the SARIMA model was a good model fitting degree in decision-making for agricultural irrigation.

Valipour (2015) studied the long-term runoff using SARIMA and ARIMA models in the United States. The ability of SARIMA and ARIMA models was investigated for the long-term forecasting in the United States. The data ranged from 1901 to 2010. The results showed that the accuracy of the SARIMA model was better than that of the ARIMA model. It was then recommended, as part of research in future, to improve the accuracy of SARIMA models.

Kaur and Rakshit (2019) used SARIMA and Periodic Autoregressive (PAR) time series models for forecasting and analysing rainfall data of Punjab, India. The two models were employed to analyse rainfall and after were compared. The results showed that the PAR model was the best model amongst the two.

Hu et al. (2007) compared time series Poisson regression model with the SARIMA model on weather variables and cryptosporidiosis in Australia. The data ranged from 1996 to 2004. It was found that both the time series Poisson regression and SARIMA models showed that seasonal and monthly maximum temperature at a prior moving average of 1 and 3 months were significantly associated with the cryptosporidiosis disease. Model assessments indicated that the SARIMA model had better predictive ability than the Poisson regression model.

Ampaw et al. (2013) used time series modelling of rainfall in New Juaben municipality of the Eastern region of Ghana. The study sought to develop a time series model for rainfall patterns and its predictions. The Box and Jenkins method was employed. The $SARIMA(0,0,1)(1,1,1)_{12}$ was the adequate model that showed that the rainfall of the following year was not different from the

rainfall of the previous years.

Akpanta et al. (2015) used the SARIMA method to model the frequency of the monthly rainfall in Umuahia, Abia State in Nigeria. The data from the National Root Crops Research Institute Umudike in Nigeria ranged from 1996 to 2011. It was found that the $SARIMA(0,0,0)(0,1,1)_{12}$ is a new model that best fit the data, and it was employed to make forecasts. The comparison of the actual and observed frequencies from July to December 2011 was done with their corresponding forecast values, and a t-test of significance showed no significant difference.

Mahsin (2012) modelled rainfall in Dhaka division of Bangladesh using time series analysis. The Box and Jenkins methodology was used to build the seasonal ARIMA model for the rainfall data from the Dhaka station. The period of the data ranged from 1981 to 2010, leading to 354 readings. The $ARIMA(0,0,1)$ $(0,1,1)_{12}$ model was found to be the adequate model for the time series data. The same model was used to forecast the monthly rainfall data for the two upcoming years.

Afrifa-Yamoah et al. (2016) used SARIMA method to model and forecast monthly rainfall in the Brong Ahafo Region of Ghana. The aim was to examine the rainfall pattern and fit a suitable model for rainfall predictions in Brong Ahafo. The SARIMA method was employed, and the $SARIMA(0,0,0)(1,1,1)_{12}$ was identified as an appropriate model for predicting monthly average rainfall figures for the Brong Ahafo Region.

Subbaiah Naidu (2016) used SARIMA model for modelling and forecasting of seasonal rainfall patterns in India. The data set ranged from 1951 to 2014, and was obtained from Open Government Data. It was found that the Seasonal

ARIMA $(0,0,0)(0,1,1)_4$ model fitted the data well and the stochastic seasonal fluctuation was successfully modelled. Apart from some extreme values, the predictions based on this model indicated that the rainfall patterns in India were almost similar. It was concluded that the increasing trend is consistent and that these changes in trend reminded them to make proper strategies for planning the agro sector and water for drinking purposes.

El-Shafie et al. (2012) compared the dynamic and static neural network models for forecasting rainfall at Klang River Basin, Malaysia. The range of the data used were from 1986 to 2008 (monthly basis) and between 1997 and 2008 (weekly basis). Their results revealed that the dynamic (temporal) neural network could be suitable for modelling the temporal dimension of the rainfall pattern and provide better forecasting accuracy. The results also showed that the dynamic neural network model with one-time step input delayed for weekly basis rainfall forecasting achieved the highest accuracy level.

Iwok and Dooga (2017) used the Box-Jenkins seasonal model to fit the temperature series. Since the assumption of the model adequacy was found to be violated, the Subset Fourier series with seasonal harmonics was introduced and added to the pure seasonal component that was found to be inadequate. The combinations resulted in a mixed seasonal and subset Fourier model with seasonal harmonics, hence the mixed models were fitted to the data. The tests revealed that the model was adequate. A comparative study was also carried out and the results showed that the mixed model performed better than the pure seasonal model and the subset Fourier model.

Abdul-Aziz et al. (2013) used the SARIMA $(0,0,0)(2,1,0)_{12}$ model to forecast the amount of rainfall in Ashanti region of Ghana. It was found that rainfall in Ashanti region significantly changes over time. The periods of low variabil-

ity and extreme variability separated by periods of transition, were also found, and forecast figures for some of the months showed an increase in the rainfall for the subsequent years. It was also noted that the model with the least AIC and BIC (Bayesian information criterion) values during a tentative model test, is the best model to be used in such analysis.

Chonge et al. (2015) used time series models for modelling the rainfall pattern of Uasin Gishu county in Kenya. The univariate Box and Jenkins, ARIMA method was fitted to the Uasin Gishu rainfall data that consisted of 456 months during the period 1977-2014. It was found that the best ARIMA model that best described the rainfall data was the SARIMA $(0,0,0)(0,1,2)_{12}$ model. The same model was used to forecast average monthly rainfall data for 24 months, from January 2015 to December 2016.

Pazvakawambwa (2017) applied the time series model when forecasting Windhoek rainfall in Namibia based on secondary monthly data during 1891-2011. The descriptive summary statistics which were measures of centrality and dispersion, time series plots, and autocorrelation functions, were generated using R software. Box and Jenkin's ARIMA modelling procedure (model identification, model estimation, model validation) was used to determine the best models for the data. The forecast values suggested that, for instance, in the 2046/47 winter season, monthly rainfall point estimates are around 15mm, which technically would be higher than expected. The lower 95% confidence limits for the same winter months were zero, highlighting the possibility of no rainfall during those periods. According to the ARIMA modelling of the Windhoek rainfall, despite the seasonal and irregular fluctuations, the mean monthly rainfall levels did not suggest an upward or downward trend over the century.

Etuk and Mohamed (2014) conducted a full-length research paper on time se-

ries analysis of monthly rainfall data for the Gadaref rainfall station in Sudan, using SARIMA methods. The 480 monthly rainfall data were used, and ranged from 1971 to 2010. The SARIMA$(0,0,0)(0,0,1)_{12}$ model was found to be the best, and was used as the basis for forecasting, planning and management of rainfall in that region.

Ariyarathna and Alibuhtto (2016) used seasonal ARIMA for modelling monthly rainfall in Katunayake Region. The univariate time series SARIMA model was developed for monthly rainfall data that ranged from 2001 to 2016 (181 observations). The results showed that the SARIMA$(2,0,2)(2,0,1)_{12}$ was the best fitted model for the rainfall data identified. The model was appropriate for forecasting the monthly rainfall for the future months, for assisting in decisions and policy makers to establish priorities for water demand, storage and disaster management in the respective field for the Katunayake region.

Todoko (2013) conducted a study on time series analysis of water consumption in the Hohoe municipality of the Volta region in Ghana. The data covered water consumption from 2009 to 2012 gathered on fortnight bases. Several methods of time series models including AR, MA, ARMA, ARIMA and SARIMA were fitted to the data, and it emerged that the most adequate model for the data was the ARIMA$(2,1,2)$.

Kibunja et al. (2014) applied SARIMA method for forecasting precipitation in Mt. Kenya Region. The objective of the study was to determine the forecasted values of precipitation and the accuracy of SARIMA model in the forecasting of precipitation. The SARIMA model which was fitted to the data exhibited the least AIC and BIC values. The data was then forecasted after following the three Box-Jenkins methodologies, that is, model identification, estimation of parameters and diagnostic check. It was found that the SARIMA model was

good for forecasting precipitation in Mt. Kenya region.

Etuk and Mohamed (2014) applied the SARIMA model for forecasting monthly rainfall in Gezira Scheme, Sudan. The monthly rainfall data was obtained from the Sudan Meteorological Authority, for the period 1971 to 2010 from Medani station. The results showed that monthly rainfall in the Gezira irrigation scheme of Sudan follows a SARIMA$(0,0,0)(0,1,1)_{12}$.

Takele (2012)) conducted a statistical analysis of rainfall patterns in Dire Dawa in Eastern Ethiopia. The main objective was to analyse rainfall patterns in Dire Dawa for the period of 30 years (1982-2011). The data was sourced from Dire Dawa and adjacent stations. Methods applied to the data were: descriptive statistics, spectrum analysis, cross-spectrum analysis and the Box-Jenkins (SARIMA) methodology. The ARIMA $(5,0,0)(0,1,1)_{12}$ was found to be the best model for the data.

Kokilavani et al. (2020) emphasised the high accuracy with which SARIMA models provide forecasts. In their study, the scholars employed the average monthly rainfall data of Coimbatore, Tamil Nadu, in India from 1991-2018. They used the Ljung-Box test to check the autocorrelation property of the residuals. The best selected model was then adopted to forecast the rainfall patterns for the successive two years, i.e., 2019 and 2020. Their study showed that the SARIMA $(0,0,0)(2,0,0)_{12}$ model was appropriate for analysing and forecasting the future rainfall patterns.

In this section we have identified some of the recent time series studies on rainfall patterns using SARIMA models. These models satisfactorily describe time series that exhibit non-stationary behaviours both within and across seasons, hence preferred.

2.6 ETS State Space Approach on Rainfall Time Series

The ETS model is a time series univariate forecasting method that focuses on trend and seasonal components (Jofipasi et al., 2018). Exponential smoothing is used to smooth a time series data by allocating more importance to recent observations and less importance to observations that are located far in time. This is done by allocating unequal weights on the observations, and the method depends more on recent observations than on old observations to perform forecast. The most weight is given to the current observation, the next highest weight to the immediately preceding observation, and the weights assigned to time series observations older in time decrease exponentially (Marera, 2016; Hyndman et al., 2002).

Hyndman et al. (2008) noted that statistical models that form the basis for exponential smoothing methods are known as innovations state space models, and they allow for an objective model selection. State space models allow for estimation of both point forecasts as well as prediction intervals. They can explicitly model the forecast error distribution which was not possible in the past and had seen them labelled as ad-hoc (Akram et al., 2007; Hyndman et al., 2002).

For each exponential smoothing method in the state space models, there exists two equivalent exponential models, one whose error distribution is additive and another whose error distribution is multiplicative, and there are 30 variations of the exponential smoothing methods in total (Hyndman and Athanasopoulos, 2018, 2017; Hyndman et al., 2008).

Modelling and forecasting using the ETS state space models have been un-

dertaken for some time, for example, Drago and Massa (2019) measured and forecasted the financial advisory demand in Italy, using a Hybrid ETS-ANN model. The hybrid forecasting model was used in two stage approach, the first one was exponential smoothing state space model and the second was autoregressive neural network model.

Gardner and Acar (2019) fitted the damped trend method of exponential smoothing in Turkey. The aim was to check whether (a) the well-established forecasting methods of exponential smoothing relied on the optimal estimation of parameters; (b) the ex ante forecast accuracy of the damped trend method of exponential smoothing can be improved by optimising parameters; and the method could be fitted according to the mean absolute error criterion rather than the mean squared error commonly used in practice. It was concluded that parameter optimisation significantly improves the forecast accuracy of the damped trend and optimisation, and it also improves the accuracy of simple smoothing.

Hyndman et al. (2008) conducted a study on the admissible parameter space for exponential smoothing models in Tokyo. The aim was to discuss the admissible parameter space for some state space models including the models that underly exponential smoothing methods. The results showed that the admissible parameter space of seasonal exponential smoothing model is much larger than that of basic structural model, leading to better forecasts from the exponential smoothing model when there is rapidly changing seasonal pattern.

Crevits and Croux (2017) performed forecasting using robust exponential smoothing with damped trend and seasonal components. The aim was to provide a framework for robust exponential smoothing for a class of exponential smoothing variants that included models with damped trend and seasonal components. The robust version of the exponential smoothing framework of Hynd-

man and Khandakar (2008) was proposed. The method is outlier robust and has robust forecasting equations, robust smoothing parameter estimation and performs robust model selection. It was implemented in the R package robets. The proposed robust method was used for outlier detection. It was noted that a time series that contains outliers can be labelled as possible anomalies and investigated further.

Ramos et al. (2015) checked the performance of state space and ARIMA models for consumer retail sales forecasting in Portugal. The aim was to compare the forecasting performance of ETS and ARIMA models when applied to the monthly sales of the five categories of women footwear of the brand Foreva using data from 2007 to 2012 (64 observations). The results showed that when automatic algorithm of the overall out-of-sample forecasting performance of state space and ARIMA models evaluated via RMSE, MAE and MAPE, is similar on both one-step and multi-step forecasts, then it was concluded that state space and ARIMA produce coverage probabilities that are close to the nominal rates for both one-step and multi-step forecasts.

Bako (2014) forecasted the pelagic fish in Malaysia using ETS state space approach. Two techniques of time series analysis were used to forecast fish catch of three commercial fish species found in the Malaysian waters, which were ETS state space and SARIMA models and data ranged from 2007 to 2011. The results, according to RMSE and MAE, demonstrated that ETS models performed better for Thunnus tonggol and Rastrelliger kanagurta, while SARIMA model performed better for Dussumiera acuta. It was concluded that both models have proven successful in describing and forecasting the monthly fishery dynamics.

Jain and Mallick (2017) conducted a study of time series models ARIMA and

ETS in India. The goal was to introduce appropriate approaches which will help in improving the efficiency of weather parameters (such as Rainfall, Relative Humidity, Wind Speed, Air Temperature), which are highly non-linear and complex phenomena. The study compared ETS and ARIMA models with estimation of MAE, MSE, MAPE and RMSE. Also included were the criteria AIC and BIC values to check which model deploys the best result for weather parameters.

Yusof and Kane (2012) modelled the monthly rainfall time series data using ETS state space and SARIMA models in Malaysia. They used monthly rainfall data from 1968 to 2003. SARIMA $(1,1,2)(1,1,1)_{12}$, SARIMA $(4,0,2)(1,0,1)_{12}$ with constant and ETS state space models based on exponential smoothing, were built. The diagnostic checking confirmed the adequacy of both models for forecasting.

Ramos and Oliveira (2016) studied a procedure for identification of appropriate state space and ARIMA models based on time-series cross-validation in Portugal. A cross-validation procedure was used to identify an appropriate ARIMA model and an appropriate state space model for a time series data of the brand Foreva ranging from 2007 to 2012. The procedure was based on one-step forecasts and used different training sets each containing one more observation than the previous one. The results showed that the procedure consistently forecasted more accurately than other approaches, and the improvements in accuracy were significant.

Panigrahi and Behera (2017) studied a hybrid ETS-ANN model for time series forecasting in India. The aim was to propose a new hybrid model for time series forecasting by combining linear and nonlinear ETS models from innovation state space (ETS) with ANN. Sixteen data sets and five different models

from literature were used to evaluate the effectiveness of the proposed model. The results revealed that no model was best for all data sets.

Murat et al. (2016) used statistical modelling of agrometeorological time series by exponential smoothing in Poland. The goal was to determine the most suitable exponential smoothing models to generate forecast using data on air temperature, wind speed and precipitation time series in Jokioinen (Finland), Dikopshof (Germany), Lleida (Spain), and Lublin (Poland). The Boreal and Continental sites where better described by additive seasonal exponential smoothing, while simple exponential smoothing was a better model for Atlantic Central and Mediterranean South sites. The exponential smoothing method had proved to be very useful for air temperature, precipitation and wind speed forecasting.

Jofipasi et al. (2018) used the ETS (Error, Trend, Seasonal) model to select the best model to forecast weather in the Aceh Besar District in Indonesia. The data set that was used contained air temperature, dew point, sea level pressure, station pressure, visibility, wind speed, and sea surface temperature during 2006-2016. The ETS (M,N,A) was used to predict air temperature and sea surface temperature; the ETS (A,N,A) was used to predict dew point, sea level pressure and station pressure; the ETS (A,A,N) was applied to predict visibility; and the ETS (A,N,N) was used to predict wind speed.

Marera (2016) applied the exponential smoothing methods to weather-related data in South Africa. The three commonly used exponential smoothing methods were applied to two data sets; rainfall and temperature, which exhibited both trend and seasonality. The method was directly applied on the data without de-seasonalising it first and also applied seasonal naive method for benchmarking the performance of exponential smoothing methods. The Holt-

Winters exponential smoothing method with additive seasonality performed best for forecasting monthly rainfall data; and the simple exponential smoothing method outperformed the Holt's and Holt-Winters methods for forecasting daily temperature data.

2.7 Chapter Summary

This chapter has presented the previous work done on rainfall, results from literature show that SARIMA and ETS state space models can be used to model and forecast rainfall. In particular, Section 2.5 and Section 2.6 show that while SARIMA and ETS state space models are useful in modelling rainfall and other weather variables, there is a dearth of literature in these areas. Furthermore, there is a shortage of literature on forecasting rainfall patterns in the Limpopo Province. Thus, this study is worthwhile. Our study is also important given the agrarian nature of the Province.

Chapter 3

Research Methodology

3.1 Introduction

This chapter provides material and methods used to conduct the study. More information about the sources of data utilised for this study is also supplied. Two different methods applied in the study are described together with descriptive statistics.

Data Source

Secondary data has been used for this study. A time series of monthly rainfall data in millimetres (mm) for the period from January 1900 to December 2016 has been used. The data was obtained from the Agricultural Research Council (ARC) for the following selected stations: Macuville Agriculture, Mara Agriculture, Marnits, Groendraai, Letaba, Pietersburg Hospital and Nebo areas of the Limpopo province of South Africa.

3.2 Time Series

A time series z_t is a sequential set of data points measured typically over successive times. The measurements may be taken every hour, day, week, month or year, or at any regular interval.

3.2.1 Time Series Analysis

Time series analysis comprises methods or processes that breakdown a series into components and explainable portions that allow trends to be identified, and estimates and forecasts to be made. Basically, time series analysis attempts to understand the underlying context of the data points through the use of a model to forecast future values based on known past values.

Time Series Components

A vital step in choosing appropriate modelling and forecasting procedure is to consider the type of data patterns displayed from the time series graphs of the time plots. The sources of variation in terms of patterns in time series data can be affected by four components which are Trend, Cyclical, Seasonal or Irregular components. The trend component is the long-term pattern of a time series; the cyclical variation is a non-seasonal component which varies in recognisable cycle; seasonal variations are recognised through observing the same repeating patterns over successive periods of time; and irregular variations occur predictably at random, usually over a very short time period.

3.3 The SARIMA Model

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is the generalisation of the well-known Box-Jenkins ARIMA model which is used to accommodate a data with both seasonal and non-seasonal features. The

ARMA model is known to be a combination of the Autoregressive (AR) and Moving Average (MA) models, and it uses past information of a given series in order to predict the future. Hamilton (1994) and Pankratz (1983) noted that the AR part of the model deals with the past observations of the series, while the MA part of the series deals with the past errors. The ARIMA model, discussed in section 2.5, is applied in the situation where the series has no seasonal features and also differenced stationary, implying that the initial differencing is required for the data to be stationary (Hamilton, 1994).

The ARIMA model with its order is usually presented as ARIMA(*p,d,q*) model, where p, d and q are integers greater than or equal to zero and refer to the order of the autoregressive, differencing and moving average parts of the model, respectively. The first parameter p refers to the number of autoregressive lags, the second parameter d refers to the order of integration that makes the data stationary and the third parameter q gives the number of moving average lags.

A process $\{z_t\}$ is said to be ARIMA*(p,d,q)* if $\Delta^d\ z_t$ is described by a stationary ARIMA(*p,q*) model. Δ means differencing of z_t in d order to achieve stationarity. Generally, we write the ARIMA model as:

$$\phi(L)(1-L)^d z_t = \theta(L)\varepsilon_t;\ \ \varepsilon_t \backsim WN(0,\sigma^2), \tag{3.1}$$

where ε_t follows a white noise (WN) process with mean 0 and variance σ^2. The autoregressive operator and the moving average operator are respectively defined as:

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p \tag{3.2}$$

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q \tag{3.3}$$

$\phi(L) \neq 0$ for $|\phi| < 1$. The process $\{z_t\}$ is stationary if and only if $d = 0$, in the case where it reduces to an ARMA(*p,q*) process.

The generalisation of ARIMA model to SARIMA model happens when the series has both seasonal and non-seasonal behaviour which causes the ARIMA model to be inefficient. This is because the model might not be able to take care of the seasonal behaviour part in the series and might mislead by giving a wrong order selection for non-seasonal components. The SARIMA model is also called the multiplication seasonal autoregressive integrated moving average model and it is denoted by ARIMA(*p,d,q*)(*P,D,Q*)$_s$ (Brockwell and Davis, 2016, 1991; Brockwell et al., 1991). Its lag form is written as:

$$\phi(L)\Phi(L^S)(1-L)^d(1-L^S)^D z_t = \theta(L)\Theta(L^S)\varepsilon_t \tag{3.4}$$

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p \tag{3.5}$$

$$\Phi(L^S) = 1 - \Phi_1 L^S - \Phi_2 L^{2S} - ... - \Phi_P L^{PS} \tag{3.6}$$

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q \tag{3.7}$$

$$\Theta(L^S) = 1 + \Theta_1 L^S + \Theta_2 L^{2S} + ... + \Theta_Q L^{QS}, \tag{3.8}$$

where

p, d and *q* are the orders of non-seasonal AR, differencing and MA, respectively.

P, D and *Q* are the orders of seasonal AR, differencing and MA, respectively.

z_t represents observable time series data at period t.

ε_t represents white noise error at period t (the error term is said to be a white noise if it has the following characteristics: $E(\varepsilon_t) = 0, E({\varepsilon_t}^2) = \sigma^2$ and $E(\varepsilon_t \varepsilon_s) = 0$ for all $t \neq s$). L represents lag shift operator ($L^k z_t = z_{t-k}$).

S represents the seasonal order (e.g. $s = 4$ is for quarterly data and $s = 12$ is for monthly data).

Box-Jenkins Approach

The Box-Jenkins method is named after the statisticians George Box and Gwilym Jenkins (Box and Jenkins, 1970). The approach proposed by Box and Jenkins came to be known as the Box-Jenkins methodology to ARIMA models in 1970 and the forecasting strategy that is based on ARIMA models. The Box-Jenkins process has four stages of model building, which are: model identification, parameter estimation, model diagnostics and forecasting.

3.3.1 Model Identification

The first step in developing the Box-Jenkins model is to check if the time series is stationary and also if there is any significant seasonality that needs to be modelled (Box and Jenkins, 1970).

Stationarity Tests

Stationary tests are designed to test the level of stationarity in order to decide on the integration order d. This is done with the Augmented Dickey-Fuller (ADF), the Phillips-Perron (PP) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests.

The ADF Test

The ADF test is used to test the null hypothesis of a unit root against the alternative of stationarity and it is based on equation (3.9) (Banerjee et al., 1993). To further explain the ADF unit root test, let us look at the general AR(p) model:

$$\phi_p(L)z_t = \varepsilon_t, \tag{3.9}$$

where we test if $\phi_p(L) = 1 - \phi_1 L - ... - \phi_p L^p$ may contain a unit root. Let $\phi_{p-1}(L) = 1 - \phi_1 L - ... - \phi_{p-1} L^{p-1}$ and assume all the roots of $\phi_{p-1}(L)$ are

outside the unit circle. Thus, we can rewrite equation (3.9) as:

$$(1-L)\phi_{p-1}(L)z_t = (z_t - z_{t-1}) - \sum_{j=1}^{p-1}\phi_j(z_{t-j} - z_{t-j-1}) = \varepsilon_t. \tag{3.10}$$

Testing for unit root implies testing if $\phi = 1$ in equation (3.11)

$$z_t = \phi z_{t-1} + \sum_{j=1}^{p-1}\phi_j \Delta z_{t-j} + \varepsilon_t. \tag{3.11}$$

The null hypothesis of the Augmented Dickey-Fuller t-test is that there exists a unit root and the alternative hypothesis of non-existence of a unit root. If the null hypothesis that there exists unit root is not rejected, this means the time series is not stationary (Hu, 2008; Dickey and Fuller, 1981, 1979).

The PP Test

The Dickey–Fuller test involves fitting the regression model:

$$\Delta z_t = \rho z_{t-1} + \textit{(constant, time trend)} + \varepsilon_t \tag{3.12}$$

by ordinary least squares (OLS), but serial correlation will present a problem. To account for this, the augmented Dickey–Fuller test's regression includes lags of the first differences of z_t. The PP test involves fitting equation (3.12), and the results are used to calculate the test statistics (Perron, 1989).

$$z_t = \pi z_{t-1} + \textit{(constant, time trend)} + \varepsilon_t. \tag{3.13}$$

In equation (3.13), ε_t is I(0) and may be heteroscedastic. The PP tests correct for any serial correlation and heteroscedasticity in the errors ε_t non-parametrically by modifying the Dickey Fuller test statistics. Phillips and Perron's test statistics can be viewed as Dickey–Fuller statistics that have been made robust to serial correlation by using the Newey and West (1987) heteroscedasticity and

autocorrelation-consistent covariance matrix estimator. One advantage of the PP tests over the ADF is that the PP tests are robust to general forms of heteroscedasticity in the error term ε_t. Another advantage is that the user does not have to specify a lag length for the test regression.

The KPSS Test

The KPSS test was established by Kwiatkowski et al. (1992) and assumes the following model:

$$z_t = \xi t + r_t + \varepsilon_t, \tag{3.14}$$

where ξ is the coefficient of a simple time trend, $\varepsilon_t \sim N(0, {\sigma_\epsilon}^2)$ and r_t is a random walk, that is, $r_t = r_{t-1} + u_t$, where u_t is a white noise process with mean zero and variance ${\sigma_u}^2$.

The null hypothesis ${\sigma_u}^2 = 0$ tests that the time series is either level $(\xi = 0)$ or trend stationarity $(\xi \neq 0)$ against the alternative that it is non-stationary. The test statistic is then derived by first fitting z_t depending on only an intercept and a trend. The resulting residuals ε_t are then used to develop a consistent estimate of the variance as follows:

$$s^2(l) = T^{-1} \sum_{i=1}^{T} {e_t}^2 + 2T^{-1} \sum_{s=1}^{l} w(s,l) + \sum_{t=s+1}^{T} e_t e_{t-s}. \tag{3.15}$$

In equation (3.15), $w(s,l) = 1 - \frac{s}{1+l}$ is the Bartlett window, which guarantees that the estimated variance is non-negative. The bandwidth l is decided by Newey-West automatic with the Bartlett kernel (Schwert, 2009). The next step is to derive the partial sum series of the residuals, that is,

$$S_t = \sum_{t=1}^{T} e_t, \tag{3.16}$$

which is developed for $t = 1, ..., T$. The results in equations (3.15) and (3.16) are

then used to develop the Langrange multiplier based on KPSS test statistics

$$\eta_\mu = \frac{\eta_\mu}{s^2(l)} = T^{-2}\frac{\sum {S_t}^2}{s^2(l)}, \tag{3.17}$$

where the critical values can be found in Kwiatkowski et al. (1992).

Seasonality Test

The seasonal frequencies can be tested to determine if the seasonal behaviour in the data is deterministic or stochastic. Usually one of the common approaches adopted is the Hylleberg-Engle-Granger-Yoo (HEGY) test (Hylleberg, 1995).

The HEGY Test

The HEGY test is an extension on the theory from the Dickey-Fuller tests to test for seasonal unit roots, and it was established by Hylleberg et al. (1990), who developed a test for quarterly data. Franses (1991) later created an extension of monthly data to implement HEGY test. The first thing is to present the seasonal difference operator Δ_s, where it follows that there should be $s = 12$ roots on the unit circle for monthly data, then this can be described by this equation:

$$\begin{aligned}\Delta_s = (1-L^{12}) = (1-L)(1+L)(1-iL)(1+iL)\times[1+(\sqrt{3}+i)L/2][1+(\sqrt{3}-i)L/2]\times \\ [1-(\sqrt{3}+i)L/2][1-(\sqrt{3}-i)L/2]\times[1+(\sqrt{3}+i)L/2][1-(\sqrt{3}-i)L/2]\times \\ [1-(\sqrt{3}+i)L/2][1+(\sqrt{3}-i)L/2].\end{aligned} \tag{3.18}$$

In equation (3.18), L is the lag shift operator and all are polynomials, except $(1-L)$ and they are all connected to the seasonal unit roots, then the test is

based on equation (3.19):

$$\varphi^*(L)z_{8,t} = \pi_1 z_{1,t-1} + \pi_2 z_{2,t-1} + \pi_3 z_{3,t-1} + \pi_4 z_{3,t-2}\pi_5 z_{4,t-1} + \pi_6 z_4, t-2+$$
$$\pi_7 z_{5,t-1} + \pi_8 z_{5,t-2} + \pi_9 z_{6,t-1} + \pi_{10} z_{6,t-2} + \pi_{11} z_{7,t-1} + \pi_{12} z_{7,t-1} + \mu_t + \varepsilon_t, \quad (3.19)$$

where μ_t is the deterministic and is specified to a constant, seasonal dummies and trend, ε_t is a white noice and $\varphi^*(L)$ is a polynomial function of L and the significance of the parameter π_i which is of interest, and the $z's$ are specified as lagged combinations of the observed time series process z_t as follows:

$$\begin{aligned}
z_{1,t} &= (1+L)(1+L^2)(1+L^4+L^8)z_t, \\
z_{2,t} &= -(1-L)(1+L^2)(1+L^4+L^8)z_t, \\
z_{3,t} &= -(1-L^2)(1+L^4+L^8)z_t, \\
z_{4,t} &= -(1-L^4)(1-\sqrt{3}L+L^2)(1+L^2+L^4)z_t, \\
z_{5,t} &= -(1-L^4)(1+\sqrt{3}L+L^2)(1+L^2+L^4)z_t, \\
z_{6,t} &= -(1-L^4)(1-L^2+L^4)(1-L+L^2)z_t, \\
z_{7,t} &= -(1-L^4)(1-L^2+L^4)(1+L+L^2)z_t, \\
z_{8,t} &= -(1-L^{12})z_t.
\end{aligned}$$

The ordinary least squares with the focus on the estimates of the π's is used to estimate the model in equation (3.19) for the specification of μ_t being considered relevant. For $\hat{\pi_i}$ with $i > 2$, the seasonal unit root at a specific frequency is only present when it happens for joined pairs of parameters. The F-test is used to test the joint two-sided null hypothesis of a unit root on the connected pairs, $(\hat{\pi_3}, \hat{\pi_4}), (\hat{\pi_5}, \hat{\pi_6}), (\hat{\pi_7}, \hat{\pi_8}), (\hat{\pi_9}, \hat{\pi_{10}})and(\hat{\pi_{11}}, \hat{\pi_{12}})$ against the alternative that there are seasonal unit roots. Then the t-test is used on $\hat{\pi_1}$ and $\hat{\pi_2}$ to test the null hypothesis of a unit root for $\hat{\pi_1}$ and $\hat{\pi_2}$ against the alternative of seasonal unit root.

Correlogram (ACF and PACF)

The correlogram is used to check the randomness of the data. The Box-Jenkins states that the autocorrelation function (ACF) and the partial autocorrelation function (PACF) can be used to identify the orders p, q, P, Q for a SARIMA model (Box et al., 2008). The ACF at lag τ is given by (Harvey, 1990):

$$r_\tau = c_\tau / c_0, \tag{3.20}$$

where c_τ is the following autocovariance function

$$c_\tau = T^{-1} \sum_{t=\tau+1}^{T} (z_t - \bar{z})(z_{t-\tau} - \bar{z}), \quad \tau = 1, 2, 3, .. \tag{3.21}$$

and c_0 is the variance derived by

$$c_0 = T^{-1} \sum_{t=1}^{T} (z_t - \bar{z})^2 \tag{3.22}$$

for T observations of the process z_t with the sample mean $\bar{z}$ (Harvey, 1990). The PACF is the autocorrelation between z_t and its lagged process z_{t+k}, while omitting all the autocorrelations ranging from z_{t+1} and z_{t+k-1}, and these are estimated with the use of derivative autocorrelation r_j, where firstly, we need the following function (Box and Jenkins, 1976):

$$r_j = \hat{\phi}_{k1} r_{j-1} + \hat{\phi}_{k2} r_{j-2} + ... + \hat{\phi}_{k(k-1)} r_{j-k+1} + \hat{\phi}_{kk} r_{j-k} \text{ for } j = 1, 2, ..., k. \tag{3.23}$$

The target is to extract the parameters $\hat{\phi}_{11}, \hat{\phi}_{22}, ..., \hat{\phi}_{kk}$, where $\hat{\phi}_{jj}$ is partial autocorrelation for lag j.

A correlogram consists of plots of the sample ACF and the PACF, both against the time lags τ. For complete randomness, the values at all lags should be

zero. The most essential part of model identification procedure is to check the significant lags, meaning the ones lying outside the interval $\pm 2/\sqrt{T}$. Hamilton (1994) further explained that the ACF and PACF are used to find the appropriate orders for SARIMA model by using the results in Table 3.1. It is assumed that the differencing implied by d and D have been done to the time series, meaning that the SARIMA$(p,d,q) \times (P,D,Q)_s$ of z_t has taken its representation as a SARIMA$(P,Q) \times (P.Q)_s$ of the differenced series Box and Jenkins (1976).

Table 3.1: ACF and PACF for identifying the orders of SARIMA$(p,q) \times (P,Q)_s$, only positive lags are of interest

	ACF	**PACF**
AR(p)	Exponentially decreasing or damped sine wave	Spikes to lag p then zero
MA(q)	Spikes to lag q then zero	Exponentially decreasing or damped sine wave
ARMA(p,q)	Exponentially decreasing or damped sine wave after $q-p$ lags	Exponentially decreasing or damped sine wave after $p-q$ lags
SAR$(P)_s$	Exponentially decreasing or damped sine wave for all lags times s	Spikes for lag P_s then zero
SMA$(Q)_s$	Spikes for lag Q_s then zero	Exponentially decreasing or damped sine wave all lags times s
SARMA$(P,Q)_s$	Exponentially decreasing or damped sine wave for all lags times s after lags$(Q-P)_s$	Exponentially decreasing or damped sine wave for all lags times s after lags$(P-Q)_s$

Goodness of Fit

For a given series, several possible models can be obtained from the ACF and PACF plots. With the ACF and PACF plots, the possible models can be obtained for the data, but it does not give the final model. The penalty function statistic such as Akaike Information Criterion (AIC or AIC$_c$) or Bayesian Information

Criterion (BIC) can be used in order to select the best model among the possible models (Akaike, 1974). The AIC, AIC$_c$ and BIC are the measures of goodness of fit for an estimated statistical model (Sakamoto et al., 1986). Given a data set, several models can be selected according to their AIC, AIC$_c$ or BIC with the one having the lowest information criterion value being the best. The criterion tries to determine the model that best describes the data with a minimum of free parameters, but it also includes a penalty that has an increasing function of the number of the estimated parameters. The penalty dejects over fitting. The general formulae of AIC, AIC$_c$ and BIC are given in equations (3.24), (3.25) and (3.26), respectively (Hurvich and Tsai, 1989; Akaike, 1974).

$$AIC = 2k - 2log(L) \;\; or \;\; 2k + nlog(\frac{RSS}{n}) \tag{3.24}$$

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1} \tag{3.25}$$

$$BIC = -2log(L) + klog(n) \;\; or \;\; log(\sigma_e^2) + \frac{k}{n}log(n), \tag{3.26}$$

where

$k =$ the number of parameters in the statistical models, $(p + q + P + Q + 1)$.

$L =$ the maximised value of the likelihood function for the estimation.

$RSS =$ the residual sum of squares of the estimated model.

$n =$ the sample size.

$\sigma_e^2 =$ the error variance.

This information criterion judges a model by how close its fitted values tend to be, to the true values looking at certain expected value. The information criterion value assigned to a model is only intended to rank the other competing models and also tell which one is the best amongst other alternatives. If two or more different models have the same AIC or BIC values, then the principles of parsimony can also be implemented in order to choose the best model. The principle states that a model with small parameters is usually better than the

complex model and some forecast accuracy test amongst the competing models can also help in deciding on the best model (Burnham and Anderson, 1998).

3.3.2 Parameter Estimation

The second step of the Box-Jenkins method for SARIMA modelling is model estimation. The method chosen in this dissertation is the maximum likelihood method. It is first assumed that the SARIMA model with parameters θ has been identified. Then it is assumed that the number of observations must be at least 50 and preferably 100 for efficient estimation (Box and Jenkins, 1976).

The first step of the maximum likelihood estimation is to specify the probability density function implied by the chosen model. It is also assumed that the error term of the model is distributed as Gaussian white noise. The estimation procedure is then performed in two steps. Firstly, the likelihood function is derived and secondly, the value of the parameter vector θ is specified to maximise the value of the function. The maximum likelihood estimate $\hat{\theta}$ is then interpreted as having the value that maximises the probability (Hamilton, 1994).

3.3.3 Model Diagnostics Residuals

The Ljung-Box test

Developed in 1978, the Ljung-Box test is a diagnostic tool used to test for the lack of fit on a time series model (Ljung and Box, 1978). The Ljung-Box test is also called the modified Box-Pierce statistic. It is the function of the accumulated sample autocorrelations, r_j, up to any specified time lag m. As a function of m, the Ljung-Box is determined as follows (Ljung and Box, 1978):

$$Q(m) = T(T+2)\sum_{j=1}^{m}\frac{r_j^2}{T-j}, \tag{3.27}$$

where

r_j =the sample autocorrelation at lag j

T =the sample size

m =the number of time lags included in the test.

The Ljung-Box test, tests the following hypothesis:

H_0: The data are independently distributed (i.e. the correlations in the population from which the sample is taken are 0, so that any observed correlations in data result from randomness of the sampling process).

H_1: The data are not independently distributed.

The choice of a plausible model depends on its p-value for the modified Box-Pierce if it is well above 0.05, indicating non-significance, meaning the bigger the p-value, the better the model.

The ARCH-LM test

The ARCH-LM test is used to check for conditional heteroscedasticity of the squared residuals a_t^2 of a given model. Consider the following linear regression model (Engle, 1982):

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + ... + \alpha_m a_{t-m}^2 + \varepsilon_t, \quad t = m+1, ..., T, \tag{3.28}$$

where ε_t implies the error term, T is the sample size and m is the prespecified positive integer. The conditional heteroscedasticity, which is also known as Arch effect, is the Lagrange Multiplier test, and it is equivalent to the usual F statistic for testing $\alpha_i = 0 \quad (i = 1, ..., m)$ from the Equation (3.28). The null hypothesis of no Arch effect in the squared residuals $(i.e., \quad \alpha_1 = ... = \alpha_m = 0)$ is checked by the test. Let $SSR_0 = \sum_{t=m+1}^{T} (a_t^2 - \bar{\omega})^2$, where $\bar{\omega} = (\frac{1}{T}) \sum_{t=1}^{T} a_t^2$ is the sample mean of a_t^2, and let $SSR_1 = \sum_{t=m+1}^{T} \hat{e}_t{}^2$, where $\hat{e}_t{}^2$ is the least squares

residual of the prior linear regression. The F statistic is given by (Tsay, 2005):

$$F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T - 2M - 1)}. \tag{3.29}$$

The null hypothesis of no arch effect is rejected if the p-value associated with F is small (p-value $< \alpha$). When the null hypothesis is satisfied, F is asymptotically $\chi^2(m)$ distributed with m degrees of freedom.

The Jargue-Bera test

The Jargue-Bera test is used to test the normality of the residuals. Then the null hypothesis of the test is normality, against the alternative of non normality (Jarque and Bera, 1980).

For a time series process z_t to follow the normal distribution with mean μ and variance σ^2, i.e., for $z_t \sim N(\mu, \sigma^2)$, it should have the following probability density function (Hamilton, 1994):

$$f(z_t) = \frac{1}{\sqrt{2\pi\sigma}} exp[\frac{-(z_t - \mu)^2}{2\sigma^2}]. \tag{3.30}$$

Furthermore, it is needed that the centred moments of odds order be equal to zero, that is:

$$E(z_t - \mu)^r = o \text{ for } r = 1, 2, 5, ... \tag{3.31}$$

and the centered fourth moment:

$$E(z_t - \mu)^4 = 3\sigma^4. \tag{3.32}$$

The two important factors while constructing the Jargue-Bera normality test, is the skewness, which is the standardised third moment, which if positive decides that it is more likely to observe a bigger number compared to the mean.

The skewness is defined as follows:

$$S = \frac{E(z_t - \mu)^3}{[Var(z_t)]^{3/2}}, \tag{3.33}$$

and the kurtosis, which is the standardised fourth moment. It measures the relationship between the amount of the density close to the mean and in the tails. The kurtosis is defined by the following equation:

$$K = \frac{E(z_t - \mu)^4}{[Var(z_t)]^2}. \tag{3.34}$$

If K is bigger than 3, then we have comparably more density in the tails compared to the normal distribution.

The statistic that is used for Jargue-Bera is as follows:

$$JB = T(\frac{S^2}{6} + \frac{K^2 - 3}{24}), \tag{3.35}$$

where T is the number of observations, S is the skewness derived from equation (3.33) and K is the kurtosis in equation (3.34). Schwert (2009) declared that the Jargue-Bera statistic is assumed to be distributed as $\chi^2_{1-\aleph}(2)$ variable and that the null hypothesis of normality is rejected for large values of the statistic.

3.3.4 Forecasting from SARIMA Models

Forecasting is the last step in Box-Jenkins model building approach. After the model has gone through all the steps of diagnostics test, it then becomes adequate for forecasting. Given a Seasonal ARIMA $(0, 1, 1)(1, 0, 1)_{12}$ model, we

can forecast the next step which is given by (Cryer and Chan, 2008):

$$z_t - z_{t-1} = \Phi(z_{t-12} - z_{t-13}) + \varepsilon_t - \theta\varepsilon_{t-1} - \Theta\varepsilon_{t-12} + \theta\Theta\varepsilon_{t-13} \tag{3.36}$$

$$z_t = z_{t-1} + \Phi z_{t-12} - \Phi z_{t-13} + \varepsilon_t - \theta\varepsilon_{t-1} - \Theta\varepsilon_{t-12} + \theta\Theta\varepsilon_{t-13}. \tag{3.37}$$

Then the one step ahead forecast from the origin t in equation (3.37) is given by the equation (3.38) and equation (3.19) as follows:

$$\hat{z}_{t+1} = z_t + \Phi z_{t-11} - \Phi z_{t-12} - \theta\varepsilon_t - \Theta\varepsilon_{t-11} + \theta\Theta\varepsilon_{t-12} \tag{3.38}$$

$$\hat{z}_{t+2} = \hat{z}_{t=1} + \Phi z_{t-10} - \Phi z_{t-11} - \Theta\varepsilon_{t-10} + \theta\Theta\varepsilon_{t-11}. \tag{3.39}$$

The noise terms $\varepsilon_{13}, \varepsilon_{12}, \varepsilon_{11}, \varepsilon_{10}, ..., \varepsilon_1$ (as residuals) will enter into the forecast for lead time $l = 1, 2, ..., 13,$ but for $l > 13$ the autoregressive part of the model takes over and we have:

$$\hat{z}_{t+l} = \hat{z}_{t=l-1} + \Phi z_{t-12} - \Phi z_{t-13} \quad for \quad l > 13. \tag{3.40}$$

3.4 Exponential Smoothing (ETS) State Space Models

Exponential smoothing methods have been used with success to generate easily reliable forecasts for a long time in time series, i.e., since the 1950 (Gardner, 1985). Although it has been used for a long time there has not been a well developed modelling framework.

The forecasts for the exponential smoothing method 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 ob-

servations. The name exponential smoothing reflects the fact that the weights decrease exponentially as the observations get older (Hyndman et al., 2008).

The most common representation of this method is the component form. This component form represents a forecast equation and a smoothing equation for each of the components included in the method. The components that might be included are level, trend and seasonal components (Ramos and Oliveira, 2016).

3.4.1 Classification of Exponential Smoothing

In exponential smoothing, we always start with the trend component which is a combination of a level (l) and a growth term (b). Level and growth terms can be combined in many ways to give five future trend types. Let T_h denote the forecast trend over the next h time periods and let ϕ denote a damping parameter ($0 < \phi < 1$). The five trend types or growth patterns are given defined as follows (Hyndman et al., 2008):

$$\begin{aligned}
\text{None} &: T_h = l \\
\text{Additive} &: T_h = l + bh \\
\text{Additive damped} &: T_h = l + (\phi + \phi^2 + ... + \phi^h)b \\
\text{Multiplicative} &: T_h = lb^h \\
\text{Multiplicative damped} &: T_h = lb(\phi + \phi^2 + ... + \phi^h).
\end{aligned}$$

A damped trend method is fitted when there is a trend in the time series, but is believed that the growth rate at the end of the historical data is only likely to continue for a short time into the future.

The equations for damped trend do what the name indicates: dampen the trend

as the length of the forecast horizon increases and this improves the forecast accuracy. After choosing the trend component, we then can introduce the seasonal component which is either additive or multiplicative. Lastly, we can include the error, either additively or multiplicatively, then the nature of the error component has often been ignored. If the error component is ignored, then we have the 15 exponential smoothing methods given in Table 3.2.

Table 3.2: Taxonomy of exponential smoothing

Trend Component	**Seasonal Component**		
	N (None)	**A** (Additive)	**PACF** (Multiplicative)
N (None)	N, N	N, A	N, M
A (Additive)	A, N	A, A	A, M
$\mathbf{A}_d$ (Additive damped)	A_d, N	A_d, A	A_d, M
M (Multiplicative)	M, N	M, A	M, M
$\mathbf{M}_d$ (Multiplicative damped)	M_d, N	M_d, A	M_d, M

This classification of methods was originated by Pegels (1969) Taxonomy and it was later extended by Gardner (1985), and then modified by Hyndman et al. (2002) and extended again by Taylor (2003) whereby giving the fifteen methods from Table 3.2. Cell N,N describes the simple exponential smoothing method, with no trend component and no seasonal component, cell A,N describes the Holt's linear method, cell A_d,N describes the damped trend method, cell A,A is the Holt-Winter's additive method and cell A,M is the Holt-Winter's multiplicative method. Then the other cells correspond to less commonly used analogous methods.

Best-known methods of point forecast

The simple introduction for some of the well-known exponential smoothing methods which are simple exponential smoothing method (N,N), Holt's linear method (A,N), damped trend method (A_d,N) and the Holt-Winters' seasonal method (A,A and A,M). The observed time series is denoted by $z_1, z_2, ..., z_n$. A

forecast of z_{t+h} based on all the data up to time t is denoted by $\hat{z}_{t+h|t}$. Then for one-step forecasts, we use the simpler notation $\hat{z}_{t+1} \equiv \hat{z}_{t+h|t}$.

3.4.1.1 Simple Exponential Smoothing Method (N,N)

The next forecast value of our time series z_t, is denoted by $\hat{z}_t$. When the observation z_t is available, then the forecast error is found to be $z_t - \hat{z}_t$. The method of simple exponential smoothing, also known as single exponential smoothing due to Brown's work in the early 1950's, takes the forecast for the previous periods and adjusts it using forecast error. The forecast for the next period is given as follows (Hyndman et al., 2008):

$$\hat{z}_{t+1} = \hat{z}_t + \alpha(z_t - \hat{z}_t), \tag{3.41}$$

where α is a constant between 0 and 1. When α has a value close to 1, then the new forecast will include a substantial adjustment for the error in the previous forecast and when α is closer to 0, then the new forecast will include a minor adjustment. Another way of writing equation (3.41) is:

$$\hat{z}_{t+1} = \alpha z_t + (1 - \alpha)\hat{z}_t. \tag{3.42}$$

Forecast $\hat{z}_{t+1}$ is based on the weighting of the most recent observation z_t with a weight value of α and weighting the most recent forecast $\hat{z}_t$ with the weight of 1-α. The interpretation is a weighted average of the most recent forecast and the most recent observations. These implications can be seen more easily if equation (3.42) is expanded by replacing $\hat{z}_t$ with its components as follows:

$$\begin{aligned}\hat{z}_{t+1} &= \alpha z_t + (1 - \alpha)[\alpha z_{t-1} + (1 - \alpha)\hat{z}_{t-1}] \\ &= \alpha z_t + \alpha(1 - \alpha)z_{t-1} + (1 - \alpha)^2\hat{z}_{t-1}.\end{aligned}$$

This substitution process is repeated by replacing $\hat{z}_{t-1}$ with its components, $\hat{z}_{t-2}$ and so on, giving the following result:

$$\begin{aligned}\hat{z}_{t+1} = \alpha z_t + \alpha(1-\alpha)z_{t-1} + \alpha(1-\alpha)^2 z_{t-2} + \alpha(1-\alpha)^3 z_{t-3} \\ +\alpha(1-\alpha)^4 z_{t-4} + ... + \alpha(1-\alpha)^{t-1} z_1 + (1-\alpha)^t \hat{z}_1, \end{aligned} \tag{3.43}$$

where $\hat{z}_{t+1}$ represents a weighted moving average of all past observations with the weights that decrease exponentially, hence the name exponential smoothing. It has been noted that the weight of $\hat{z}_t$ might be quite large when α is small and the time series is relatively short. For longer range forecasts, it is assumed that the forecast function is flat, that is:

$$\hat{z}_{t+h|t} = \hat{z}_{t+1}, \quad h = 2, 3, ...$$

The flat forecast function is used because the simple exponential smoothing works best for data that have no trend, seasonality, or other patterns (Hyndman et al., 2008).

3.4.1.2 Holt's Linear Method (A,N)

Holt (1957) extended simple exponential smoothing to a linear exponential smoothing method so that it can allow forecasting of data that has trends. Then the forecast for this method is computed using two smoothing constants α and β^* (with values between 0 and 1) and the following three equations:

$$\text{Level}: \ l_t = \alpha z_t + (1-\alpha)(l_{t-1} + b_{t-1}), \tag{3.44}$$

$$\text{Growth}: \ b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}, \tag{3.45}$$

$$\text{Forecast}: \ \hat{z}_{t-h|t} = l_t + b_t h, \tag{3.46}$$

where l_t denotes an estimate of the level of the series at time t and b_t denotes an estimate of the slope of the series at time t. This method is also called the double exponential smoothing, in the special case where $\alpha = \beta^*$. The method is equivalent to Brown's double exponential smoothing method (Brown, 1959).

3.4.1.3 Damped Trend Method (A_d,N)

The damped trend method was proposed by Gardner and McKenzie (1985) as a modification of Holt's linear method so that it can allow the damping of trends. The method used the same parameterisation as in Gardner and McKenzie (1985) which is slightly different from the parameterisation proposed by Hyndman et al. (2002). It makes no difference to the value of the forecasts. The equations for this method are as follows:

$$\text{Level}: \ l_t = \alpha z_t + (1-\alpha)(l_{t-1} + \phi b_{t-1}), \tag{3.47}$$

$$\text{Growth}: \ b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)\phi b_{t-1}, \tag{3.48}$$

$$\text{Forecast}: \ \hat{z}_{t+h|t} = l_t + (\phi + \phi^2 + ... + \phi^h) b_t. \tag{3.49}$$

The growth for one-step forecast of z_{t+1} is ϕb_t and the growth is dampened by a factor of ϕ for each additional future time period. Then if $\phi = 1$, this method gives the same forecasts as Holt's linear method. For $0 < \phi < 1$, as $h \to \infty$ the forecasts approach an asymptote given by $l_t + \phi b_t/(1-\phi)$. We often restrict $\phi > 0$ to avoid a negative coefficient being applied to b_{t-1} in equation (3.49) and $\phi \leq 1$ to avoid b_t increasing exponentially (Hyndman et al., 2008).

3.4.1.4 Holt-Winters' Trend and Seasonality method (A,A and A,M)

If the data have no trend and seasonality patterns, then simple exponential smoothing is the proper method to use. If the data exhibit a linear trend, then the Holt's linear method is appropriate. But if the data is seasonal, then the simple exponential smoothing and Holt's linear method on their own cannot

handle the problem well.

The Holt's method was extended by Winters (1960) to capture seasonality. Holt's Winters' method is based on three smoothing equations: one for the level, one for trend and one for seasonality. This method is similar to Holt's method with one additional equation to deal with seasonality. There are two different Holt-Winters' methods, depending on whether seasonality is modelled in an additive way or multiplicative way.

Multiplicative seasonality (A,M method)

These methods are slightly different from the usual Holt-Winters equation like those in the Makridakis, Wheelwright and Hyndman (1998). This modification was proposed by HKSG (Hyndman, Koehler, Snyder and Grose) to make the state space formulation simpler. The basic equations of Holt-Winters multiplicative method are as follows:

$$\text{Level}: \; l_t = \alpha \frac{z_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{3.50}$$

$$\text{Trend}: \; b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1} \tag{3.51}$$

$$\text{Seasonal}: \; s_t = \gamma z_t/(l_{t-1} + b_{t-1}) + (1-\gamma)s_{t-m} \tag{3.52}$$

$$\text{Forecast}: \; \hat{z}_{t+h|t} = (l_t + b_t h)s_{t-m+h_m^+}, \tag{3.53}$$

where m is the length of seasonality, l_t represents the level of the series, b_t denotes the growth, s_t is the seasonal component, $\hat{z}_{t+h|t}$ is the forecast for h periods ahead and $h_m^+ = [(h-1)mod \;\; m] + 1$ and the parameters ($\alpha, \;\; \beta^*$and $\;\gamma$) are usually restricted to lie between 0 and 1.

Additive seasonality (A,A method)

The seasonal component in Holt-Winters' method might also be treated additively, although this is less common. The basic equations for Holt-Winters' additive methods are as follows:

$$\text{Level}: \; l_t = \alpha(z_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{3.54}$$

$$\text{Trend}: \; b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1} \tag{3.55}$$

$$\text{Seasonal}: \; s_t = \gamma(z_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \tag{3.56}$$

$$\text{Forecast}: \; \hat{z}_{t+h|t} = l_t + b_t h + s_{t-m+h_m^+}. \tag{3.57}$$

Equation (3.55) is identical to equation (3.51). The only difference in these equations, is that the seasonal indices in equation (3.56), are now added and subtracted instead of taking products and ratios.

3.4.2 Innovations State Space Model

In this section, we introduce the state space models, that underlie exponential smoothing methods. The statistical models that form the basis for exponential smoothing methods are known as innovations state space models. These innovations models provide a general framework for statistical modelling with exponential smoothing, by allowing the objective of the model selection. These state space models allow the estimation of both the point forecast and the prediction intervals (Hyndman and Athanasopoulos, 2018, 2017, 2014; Hyndman et al., 2008). They can model the forecast error distribution, which was not so in the past; and they are labelled as ad-hoc (Ord et al., 1997).

In the state space, each model consists of a measurement equation that describes the observed data and some transition equations that describe how the unobserved components or the states (level, trend and seasonal) change over

time, hence being referred to as the state space models.

For each method, there exists two models: one with additive errors and one with multiplicative errors. Point forecasts produced by the models are identical if they use the same smoothing parameter values. To differentiate between a model with additive errors and the one with multiplicative errors, we add a third letter to the classification of Table 3.2 and label each state space model as ETS (.,.,.) for (Error, Trend, Seasonal). This is also called Exponential Smoothing. Using the same notations as in Table 3.2, the possibilities for every component is: Error $= \{A, M\}$, Trend $= \{N, A, A_d, M, M_d\}$ and seasonal $= \{N, A, M\}$. Therefore, in total, there exist 30 state space models: 15 with additive errors and 15 with multiplicative errors (Hyndman et al., 2008, 2002).

ETS (A,N,N): Simple exponential smoothing with additive errors

The error correction form of a simple exponential smoothing is as follows:

$$l_t = l_{t-1} + \alpha e_t, \tag{3.58}$$

where $e_t = z_t - l_{t-1}$ and $\hat{z}_{t|t-1} = l_{t-1}$. Then $e_t = z_t - \hat{z}_{t|t-1}$ that represents a one-step forecast error, can be written as $z_t = l_{t-1} + e_t$.

To develop an innovations state space model from equation 3.58, we need to specify the probability distribution for e_t. In the model with the additive errors, we assume that one-step forecast errors e_t are normally distributed with a white noise, with mean 0 and variance σ^2. It's notation is $e_t = \varepsilon_t \sim \text{NID}(0, \sigma^2)$; where NID stands for normally independently distributed. Then the equation

of the model can be written as:

$$z_t = l_t + \varepsilon_t \tag{3.59}$$

$$l_t = l_{t-1} + \alpha\varepsilon_t. \tag{3.60}$$

Equation (3.59) is the measurement (or observation) equation and equation (3.60) is the state (or transition) equation. These two equations together with the statistical distribution of the errors form a fully specified statistical model. This constitutes: an innovations state space model underlying simple exponential smoothing. This term ”innovations” comes from the fact that all equations in this type of specification use the same random error process ε_t and it is also referred to as a ”single source of error” model, different to the alternative multiple source of error formulation.

The measurement equation shows the relationship between the observations and the unobserved states. The observation z_t is a linear function of the level l_{t-1}, the predictable part of z_t and the random error ε_t, is the unpredictable part of z_t and for other innovations state space models, this relationship might be non-linear.

The transition equation shows the evolution of the state through time. The smoothing parameter α governs the degree of change in successive levels. The higher the value of α, the more rapid the changes in the level; and the lower the value of α, the smoother the changes. At the lowest extreme where $\alpha = 0$, the level of the series does not change over time and at the other extreme where $\alpha = 1$, the model reduces to a random walk model, $z_t = z_{t-1} + \varepsilon_t$ (Hyndman et al., 2008).

ETS (M,N,N): Simple exponential smoothing with multiplicative errors

In a similar manner, we can specify models with multiplicative errors by writing the one-step random errors as relative errors, as follows:

$$\varepsilon_t = \frac{z_t - \hat{z}_{t|t-1}}{\hat{z}_{t|t-1}}, \tag{3.61}$$

where $\varepsilon_t \sim$NID$(0, \sigma^2)$. Substituting $\hat{z}_{t|t-1} = l_{t-1}$ gives $z_t = l_{t-1} + l_{t-1}\varepsilon_t$ and $e_t = z_t - z_{t|t-1} = l_{t-1}\varepsilon_t$.

We can write the multiplicative form of the state space model as:

$$z_t = l_{t-1}(1 + \varepsilon_t) \tag{3.62}$$

$$l_t = l_{t-1}(1 + \alpha\varepsilon_t). \tag{3.63}$$

ETS (A,A,N): Holt's linear method with additive errors

We assume that one-step forecast errors are given by $\varepsilon_t = z_t - l_{t-1} - b_{t-1} \sim$ NID $(0, \sigma^2)$. Substituting this into the error correction equations for Holt's linear method, we obtain:

$$z_t = l_{t-1} + b_{t-1} + \varepsilon_t \tag{3.64}$$

$$l_t = l_{t-1} + b_{t-1} + \alpha\varepsilon_t \tag{3.65}$$

$$b_t = b_{t-1} + \beta\varepsilon_t, \tag{3.66}$$

where for simplicity, we have set $\beta = \alpha\beta^*$.

ETS (M,A,N): Holt's linear method with multiplicative errors

We specify one-step forecast errors as relative errors such that:

$$\varepsilon_t = \frac{z_t - (l_{t-1} + b_{t-1})}{(l_{t-1} + b_{t-1})}. \tag{3.67}$$

Following the same approach as in equation (3.67), the innovations state space model underlying Holt's linear method with multiplicative errors is specified as:

$$z_t = (l_{t-1} + b_{t-1})(1 + \varepsilon_t) \tag{3.68}$$

$$l_t = (l_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t) \tag{3.69}$$

$$b_t = b_{t-1} + \beta(l_{t-1} + b_{t-1})\varepsilon_t, \tag{3.70}$$

where again $\beta = \alpha\beta^*$ and $\varepsilon_t \sim$NID$(0, \sigma^2)$.

3.4.3 Initialisation and Estimation

Initialisation

The non-linear optimisation requires some initial value, hence we use $\alpha = \beta = \gamma = 0.5$ and $\phi = 0.9$. Then the initial values of l_0, b_0 and $s_k(k = -m + 1, ..., 0)$ are obtained using the following procedure:

- Seasonal data: we compute a $2 \times m$ moving average through the first few years of data. Denote this by $\{f_t\}$, $t = m/2 + 1, m/2 + 2, ...$
- Additive seasonality: we detrend the data to get $z_t - f_t$, then multiplicative seasonality, we detrend data to get $z_{/}f_t$, and we then compute the initial seasonal indices $s_{-m+1}, ..., s_0$ by averaging the detrended data for each season over the first 3 years available (from $t = m/2+1$ to $t = 7m/2$). Then we normalise the seasonal indices, that they might add to zero for additive seasonality and to m for multiplicative seasonality.
- For seasonal data, we compute a linear trend using linear regression on the first 10 seasonally adjusted values against a time variable $t = 1, ..., 10$.
- For non-seasonal data, we compute a linear trend on the first 10 observations against a time variable $t = 1, ..., 10$. We then set l_0 to be the intercept

of the trend.

- For additive trend, we set b_0 to be the slope of the trend and for multiplicative trend, we set $b_0 = 1 + b/a$, where a denotes the intercept and b denotes the slope of the fitted trend.

These values of the initial state X_0 are then refined by estimating them as parameters along with the elements θ.

Estimation

The alternative way of estimating the parameters is to minimise the sum of squared errors and to maximise the likelihood. The likelihood is the probability of the data arising from the specified model. The large likelihood is associated with a good model.

$$L^*(\theta_t, X_0) = n\text{log}(\sum_{t=1}^{n} e_t^2/k^2(x_{t-1})) + 2\sum_{t=1}^{n} \text{log}|k(x_{t-1})|. \tag{3.71}$$

Equation (3.71) is equal to twice the negative logarithm of the likelihood function (with constant terms eliminated) conditional on the parameters $\theta = (\alpha, \beta, \gamma, \phi)'$. The initial states $x_0 = (l_0, b_0, s_0, s_{-1}, ..., s_{-m+1})'$, can be estimated by minimising L^*. Alternatively, estimates can be done by minimising the one-step MSE, minimising the one step-step MAPE and minimising the residual variance σ^2 or via some other criteria for measuring forecast error. We make the estimation by restricting the parameters to lie within the following intervals:
$0.1 \leq \alpha \leq 0.9, \quad 0.1 \leq \beta \leq 0.9, \quad 0.1 \leq \gamma \leq 0.9, \quad \beta \leq \phi \leq 1.$

Mostly α, β and γ are restricted to values in the range (0,1). Then we use a smaller range to avoid instabilities occurring. We also constrain the initial states X_0 so that the seasonal indices add to zero for additive seasonality and add to m for multiplicative seasonality (Hyndman et al., 2008).

3.4.4 Model Selection

A great advantage of the ETS statistical framework is that information criteria can be used for model selection. The AIC, AIC_c and the BIC can be used to determine which of the 30 ETS models is appropriate for a given time series. For ETS models, the Akaike Information Criterion (AIC) is defined as:

$$AIC = -2log(L) + 2k, \tag{3.72}$$

where L is the likelihood of the model, k is the total number of parameters and initial states that have been estimated including the residuals.

The AIC corrected for small sample bias (AIC_c) is defined as:

$$AIC_c = AIC + \frac{2k(k+1)}{T-k-1}. \tag{3.73}$$

The Bayesian Information Criterion (BIC) is defined as in equation (3.74):

$$BIC = AIC + k[log(T) - 2]. \tag{3.74}$$

3.4.5 Forecasting with ETS Model

We combine the preceding ideas to obtain a robust and widely applicable automatic forecasting algorithm. The steps are as follows:

- For each series, we apply the models that are appropriate for optimising the parameters of the model in each case.
- We select the best of the models according to the AIC and others.
- We produce forecasts using the best model (with optimised parameters) for as many steps ahead as required.

- To obtain prediction intervals, we use a bootstrap method by simulating 5000 future sample paths for $\{z_{n+1}, ..., z_{n+h}\}$ and finding the $\alpha/2$ and $1 - \alpha/2$ percentiles of the simulated data at each forecasting horizon. The sample paths are generated using the normal distribution for errors (parametric bootstrap) or using the resampled errors (ordinary bootstrap).

With the linear state space models, it is possible to compute analytical prediction intervals (Hyndman et al., 2002). However, this is difficult with non-linear models and so we prefer to use a simulation approach which is applicable to all the models and easy to implement (Hyndman et al., 2002).

3.5 Comparing Forecasts (Error Measures)

The forecast accuracy of the two models obtained from SARIMA and ETS State Space will be compared with some error measures which are RMSE, MSE, MAE and MAPE. The error measures are easily derived and give an indication of the model that is most fit for forecasting. This will be derived by the use of z_t which is the real value and $\hat{z}_t$ is the forecast value at time t. Here n is the total number of observations and ESS is the error sum of squares (Petris et al., 2009). The first error measure that is presented is the **Mean square error (MSE)**:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(z_t - \hat{z}_t)^2. \tag{3.75}$$

A value of zero would mean a perfect forecast and a negative value is not possible to get.

The next error measure is the **Root mean square error (RMSE)**:

$$RMSE = \sqrt{\frac{ESS}{n}}. \tag{3.76}$$

The RMSE measures the mean prediction error, with RMSE $= 0$ implying a perfect fit. The RMSE is also called the root mean square deviation (RMSD). It is frequently used to measure the difference between values predicted by a model and the values actually observed from the environment that is being modelled.

The following error measure is the **mean absolute error (MAE)**:

$$MAE = \frac{1}{n}\sum_{t=1}^{n} \mid z_t - \hat{z}_t \mid . \tag{3.77}$$

An MAE measure with the value of zero, would mean a perfect forecast and a negative value is not possible. The only difference between MSE and MAE is that the MSE places a relatively greater penalty on large forecast errors.

The last measure of error is the **mean absolute percentage error (MAPE)**:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \mid \frac{z_t - \hat{z}_t}{z_t} \mid . \tag{3.78}$$

The value of zero would mean a perfect forecast and a negative value does not exist. This measure is not defined if $z_t = 0$ for any time t. The MAPE scales the errors and puts relatively more penalty to forecast errors if the true value of the observation is small (Petris et al., 2009).

3.6 Chapter Summary

This chapter has presented the techniques, methods and steps to be applied when modelling rainfall. The data is analysed using SARIMA and ETS State Space models. Most research conducted on rainfall, relied on linear models such as ARIMA and SARIMA. These models are often incapable of modelling

non-linear behaviour, which are often related to rainfall data. This study will employ a non-linear model together with a linear model and compare their forecasting ability using MSE and RMSE, among others.

Chapter 4

Results and Discussion

4.1 Introduction

Having explored the theory of SARIMA and ETS State Space models in the previous chapters, this chapter is dedicated to fitting the two models of time series to the rainfall data for all the selected stations in the Limpopo Province. A description of the data for all the stations is given in section 4.2. Application of SARIMA and ETS State Space is given in sections 4.3 and 4.4 respectively, and the comparison of the two models is given in section 4.5. Analysis was done using R software package.

4.2 Descriptive Statistics

Chatfield (1995) noted that exploratory data analysis (EDA) is crucial to any successful statistical modelling. EDA involves an in-depth description and assessment of the quality of data. Descriptive measures such as summary statis-

tics and graphical methods are used to gain insight into the data.

Table 4.1 presents summary statistics for the rainfall quantities received over the period of January 1900 to December 2016 in Limpopo Province for all the seven rainfall stations. The time series for all these stations of rainfall comprises of 1404 monthly observations.

The estimated average monthly rainfall received in Macuville over the period is 36.3mm, and ranges from zero (no rain received at all) to 517.7mm with the standard deviation of 56.9mm. The estimated average monthly rainfall received in Mara over the period is 39.9mm, and ranges from zero (no rain received at all) to 468.1mm with the standard deviation of 53.3mm. The estimated average monthly rainfall received in Marnits over the period is 35.9mm, and ranges from zero (no rain received at all) to 412.3mm with the standard deviation of 46.9mm. The estimated average monthly rainfall received in Groendraai over the period is 49.9mm, and ranges from zero (no rain received at all) to 393.7mm with the standard deviation of 57.1mm. The estimated average monthly rainfall received in Letaba over the period is 43.7mm, and ranges from zero (no rain received at all) to 734.7mm with the standard deviation of 59.6mm. The estimated average monthly rainfall received in Pietersburg over the period is 39.8mm, and ranges from zero (no rain received at all) to 368.9mm and with the standard deviation of 48.4mm. The estimated average monthly rainfall received in Nebo over the period is 51.7mm, and ranges from zero (no rain received at all) to 756.9mm with the standard deviation of 57.1mm.

At 51.7mm, Nebo Station received the highest average monthly rainfall during the period under consideration. The second highest average amount of rainfall (49.9mm) was recorded for Groedraai Station, followed by Letaba with an average of 43.7mm. The remaining four stations received an average monthly

rainfall of below 40.0mm, with Martis Station receiving the lowest amount of rainfall at 35.9mm. The least variation in rainfall patterns is recorded for Martis with a standard deviation of 46.9, while the highest variation is observed at Letaba with a standard deviation of 59.6.

Table 4.1: Descriptive statistics for rainfall pattern in all the stations.

		Descriptive Statistic			
Weather Stations	**No. of Observations**	**Min**	**Max**	**Mean**	**Std. Dev.**
Macuville	1404.0	0.0	517.7	36.3	56.8
Mara	1404.0	0.0	468.1	39.9	53.3
Marnits	1404.0	0.0	412.3	35.9	46.9
Groendraai	1404.0	0.0	393.7	49.9	57.1
Letaba	1404.0	0.0	734.7	43.7	59.6
Pietersburg	1404.0	0.0	368.9	39.8	48.4
Nebo	1404.0	0.0	756.9	51.7	57.0

The time series plots were obtained together with the decomposition plot. From Figure 4.1, it can be seen that the data set of each station exhibits high fluctuations and there is no visible trend. The time series is stationary since the mean and variance of the series do not seem to vary with the level of the series, making additive decomposition ideal for the separation of time series components. The results of additive time series decomposition are presented in Figure 4.2.

For each station in Figure 4.2, are shown: the original time series (top); the estimated trend component (second from top); the estimated seasonal component (third from the top); and the estimated irregular component (bottom). According to the plots, the estimated trend component shows no evidence of a moving trend.

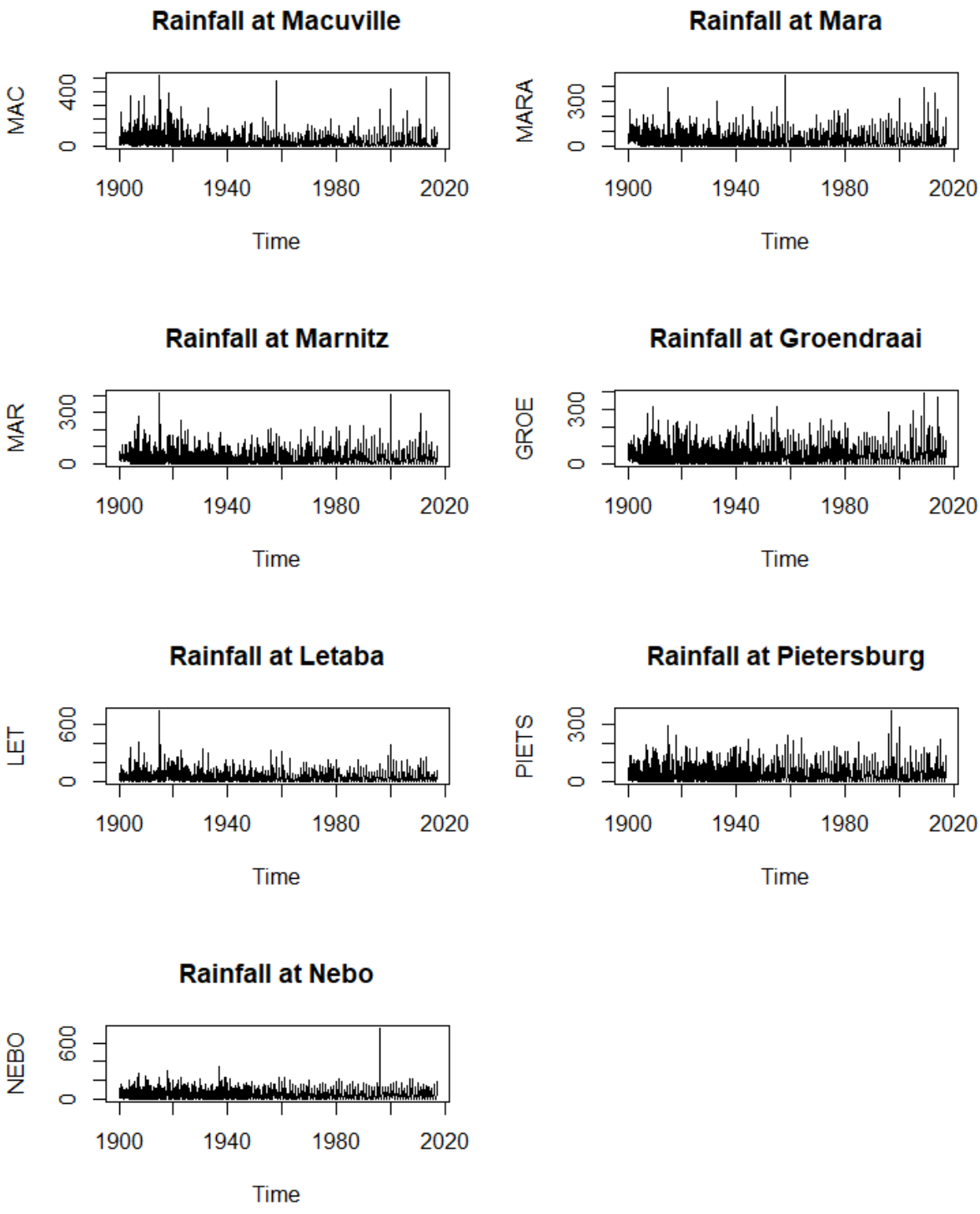


Figure 4.1: Time series plots for monthly rainfall data from January 1900 to December 2016 for all 7 stations of Limpopo Province.

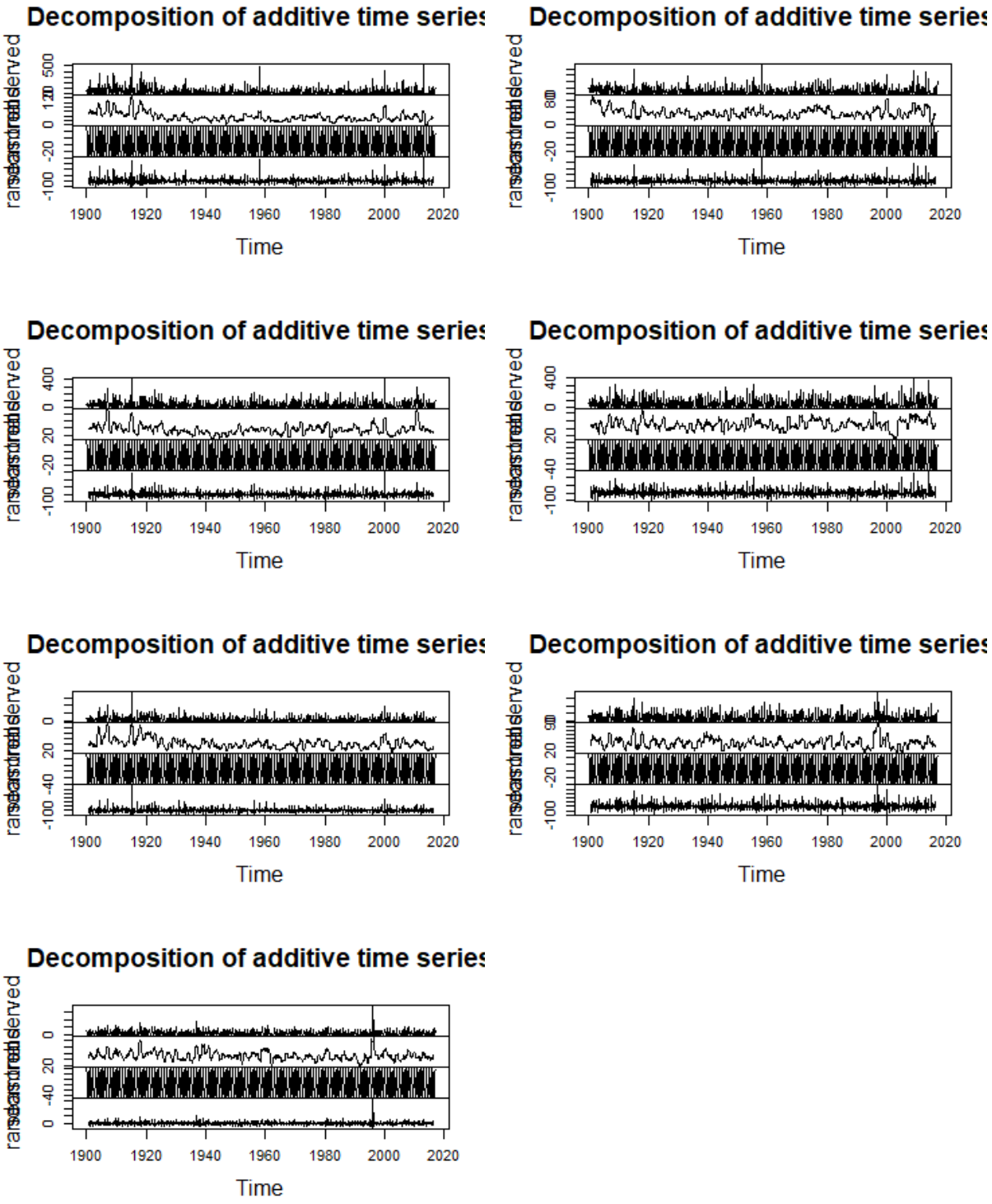


Figure 4.2: Decomposed time series plots for monthly rainfall data from January 1900 to December 2016 for all 7 stations of Limpopo Province.

4.3 The SARIMA Model

4.3.1 Model Identification and Estimation

Stationarity

The Augmented Dickey-Fuller (ADF), the Phillips-Perron (PP) and the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) tests were used to ascertain the stationarity of data for all the selected stations of the Limpopo Province. Table 4.2 reports the results of ADF, PP and KPSS in levels. Stationary simply implies that the variable has zero mean, constant variance and the residuals uncorrelated over time. Looking at Table 4.2 for all the selected stations, there is no unit root, therefore, the series is stationary.

Table 4.2: Stationary tests.

Weather Stations	**ADF**	**PP**	**KPSS**
Macuville	-8.5660(0.0100)***	-693.2400(0.0100)***	2.2024(0.0100)***
Mara	-8.5119(0.0100)***	-748.9800(0.0100)***	0.3408(0.1000)***
Marnits	-8.4210(0.0100)***	-661.0000(0.0100)***	0.2696(0.1000)***
Groendraai	-9.1641(0.0100)***	-551.3300(0.0100)***	0.1193(0.0100)***
Letaba	-8.6752(0.0100)***	-708.6300(0.0100)***	1.5405(0.0100)***
Pietersburg	-9.4013(0.0100)***	-633.7900(0.0100)***	0.0463(0.1000)***
Nebo	-9.6460(0.0100)***	-557.3800(0.0100)***	0.3138(0.1000)***

NB: *Decision was based on the ADF test which is considered as the most powerful test. *** In all the cases, the null hypothesis of the unit root is rejected. p-values are presented in brackets.*

Seasonality

Seasonality in a time series reflects a pattern of ups and downs over a fixed period of time. We next employ the HEGY test to evaluate if any of the processes has a seasonal unit root. The test is defined with an intercept and seasonal

dummies since the seasonality is assumed, but the trend is not included following what was stated for the tests level. The results of the HEGY test are presented in Table 4.3. In the HEGY test, the null hypothesis of a seasonal unit root is tested against the alternative hypothesis of a unit root. The seasonal unit root is rejected on the five percent significance level for all the stations.

The order of integration needed to make each series stationary has now been decided. The procedure of identifying the SARIMA can continue to the next step which is to determine the AR and MA orders.

Table 4.3: HEGY test for seasonality.

Weather	**HEGY test**	
	Row data	
Joint tests	**Est.**	**p-value**
Macuville	101.9656	0.0000***
Mara	84.2638	0.0000***
Marnits	84.0934	0.0000***
Groendraai	83.7622	0.0000***
Letaba	84.7157	0.0000***
Pietersburg	83.4468	0.0000***
Nebo	88.2936	0.0000***
Seasonal difference(12 for monthly)		
Joint tests	**Est.**	**p-value**
Macuville	373.4962	0.0000***
Mara	327.3673	0.0000***
Marnits	337.1201	0.0000***
Groendraai	372.4654	0.0000***
Letaba	341.8509	0.0000***
Pietersburg	337.1649	0.0000***
Nebo	374.884	0.0000***

NB: **** in all cases the null hypothesis of seasonal unit root is rejected after first differencing (12 for monthly).*

The data set from each station was divided into two parts: 90% in-sample and

10% out-of-sample. The in-simple data set was used for model fitting and the out-of-simple was used to validate the fitted models.

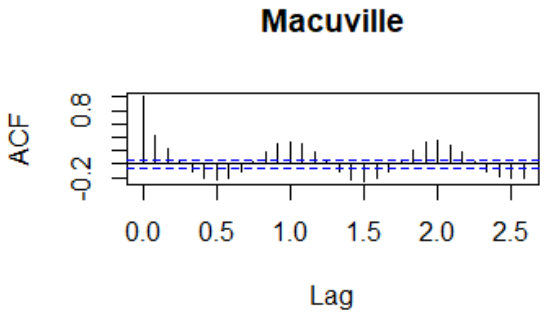
Autocorrelation functions and partial autocorrelation functions are the sample statistics utilised in identifying an appropriate ARMA model. The model identification procedures allow us to build up a model systematically, taking into account the main features exhibited in the sample ACF and sample PACF. Table 3.1 (in the previous chapter) gives the criteria for model identification given by Kendall and Ord (1990). A closer look at the ACF plot in Figure 4.4 obtained from the series up to lag 12, indicates that there is evidence of exponential decay and damped oscillations. This may suggest the presence of both AR and MA parameters, hence an ARMA process with both ordinary and seasonal terms can be considered. There is a large spike at lag 12, which suggests a seasonal parameter.

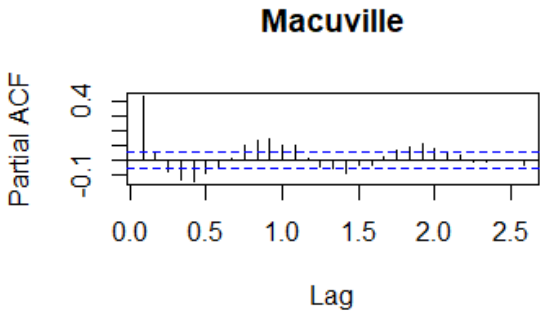
In light of a number of competing models, the criteria used to select the best fit model to the data is the AIC. A parsimonious model (one with fewer parameters) with a lower AIC is then the best fitted model. The aim is to obtain the values of p, q, d, P, Q and D of the SARIMA model, where P and Q represent the seasonal AR and MA orders respectively, and D represents the order of seasonal differencing.

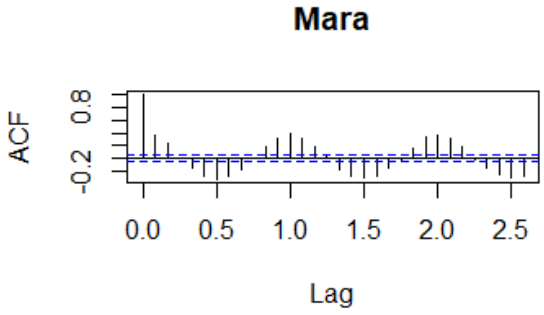
After considering many possible models in each station, the best selected model for each station are as follows:
For Macuvlle is $SARIMA(1,0,1)(1,1,1)_{12}$,
For Mara is $SARIMA(1,0,1)(1,1,1)_{12}$,
For Marnits is $SARIMA(2,0,2)(1,1,0)_{12}$,
For Groendraai is $SARIMA(2,0,1)(2,1,1)_{12}$,
For Letaba is $SARIMA(1,0,2)(1,1,2)_{12}$,

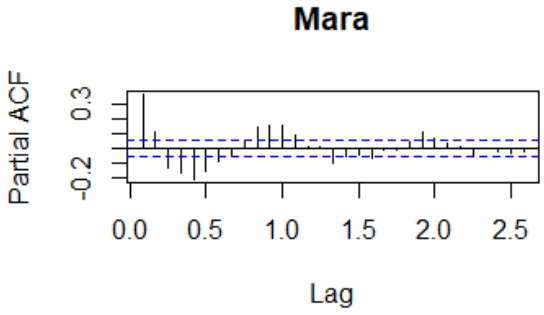
For Pietersburg is $SARIMA(1,0,0)(0,1,1)_{12}$ and
For Nebo is $SARIMA(1,0,1)(1,1,1)_{12}$.

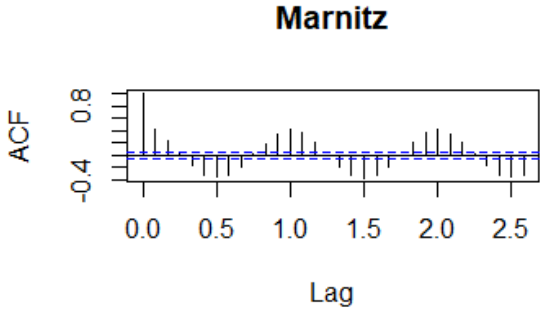
All the models had significant parameters and the lowest values of AIC, amongst others. The results are shown in Table 4.4 for all the stations, together with their parameter estimates, standard errors, the AR and MA orders, and the selected best models by the Akaike information criterion (AIC).

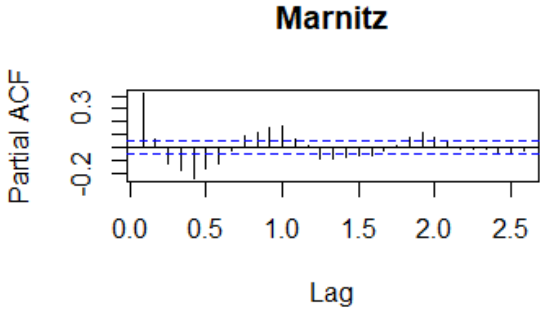

Macuville
ACF
0.8
-0.2
0.0
0.5
1.0
1.5
2.0
2.5
Lag

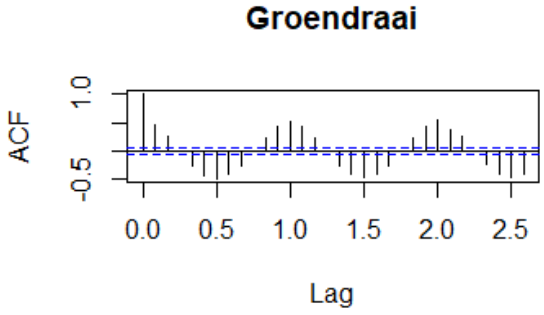

Macuville
Partial ACF
0.4
-0.1
0.0
0.5
1.0
1.5
2.0
2.5
Lag

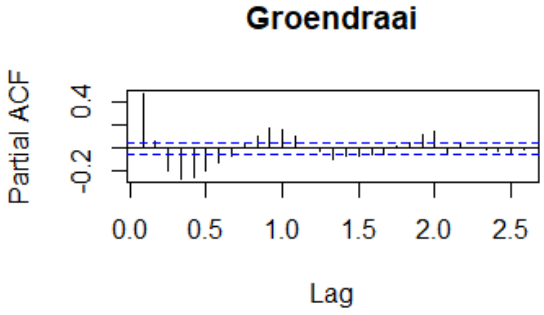

Mara
ACF
0.8
-0.2
0.0
0.5
1.0
1.5
2.0
2.5
Lag


Mara
Partial ACF
0.3
-0.2
0.0
0.5
1.0
1.5
2.0
2.5
Lag


Marnitz
ACF
0.8
-0.4
0.0
0.5
1.0
1.5
2.0
2.5
Lag


Marnitz
Partial ACF
0.3
-0.2
0.0
0.5
1.0
1.5
2.0
2.5
Lag


Groendraai
ACF
1.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
Lag


Groendraai
Partial ACF
0.4
-0.2
0.0
0.5
1.0
1.5
2.0
2.5
Lag

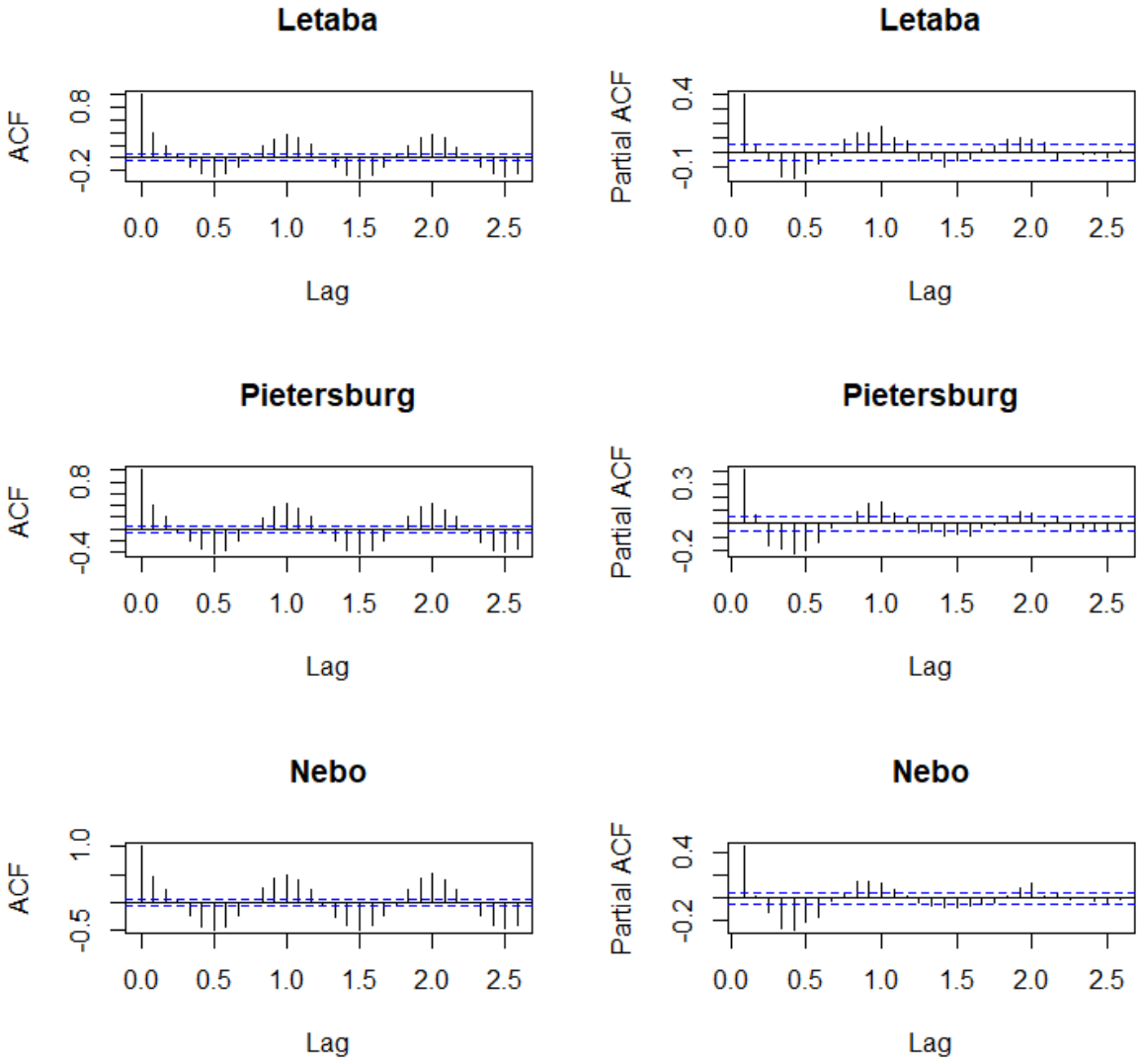


Figure 4.4: The Autocorrelation function and Partial autocorrelation functions for all the 7 stations of Limpopo Province.

Table 4.4: Parameter estimates for the fitted seasonal ARIMA models or summary results for the fitted seasonal ARIMA models

Weather Station	**Macuville**		**Mara**	
Model	$SARIMA(1,0,1)(1,1,1)_{12}$		$SARIMA(1,0,1)(1,1,1)_{12}$	
Parameter	Est.	S.E.	Est.	S.E.
$AR(1)$	0.3948	0.1205	0.7211	0.1728
$MA(1)$	-0.1822	0.1284	-0.6402	0.1920
$SAR(1)$	-0.0567	0.0298	0.0310	0.0280
$SMA(1)$	-0.9293	0.0149	-0.9934	0.0241
$SMA(2)$	-	-	-	-
Weather Station	**Marnitz**		**Groendraai**	
Model	$SARIMA(2,0,2)(1,1,0)_{12}$		$SARIMA(2,0,1)(2,1,1)_{12}$	
Parameter	Est.	S.E.	Est.	S.E.
$AR(1)$	0.0938	0.8049	0.1289	0.3596
$MA(1)$	0.0049	0.8012	-0.0313	0.3597
$AR(2)$	0.3354	0.4472	0.0650	0.0468
$MA(2)$	-0.2874	0.3636	-	-
$SAR(1)$	-0.4853	0.0234	0.0248	0.0270
$SMA(1)$	-	-	-1.0000	0.0151
$SAR(2)$	-	-	0.0990	0.0278
Weather Station	**Letaba**		**Petersburg**	
Model	$SARIMA(1,0,2)(1,1,2)_{12}$		$SARIMA(1,0,0)(0,1,1)_{12}$	
Parameter	Est.	S.E.	Est.	S.E.
$AR(1)$	0.9955	0.0060	0.0838	0.0269
$MA(1)$	-0.8607	0.0276	-0.9902	0.0151
$MA(2)$	-0.1182	0.0277	-	-
$SAR(1)$	-0.4284	0.4467	-	-
$SMA(1)$	-0.5315	0.4421	-	-
$SMA(2)$	-0.4442	0.4330	-	-
Weather Station	**Nebo**			
Model	$SARIMA(1,0,1)(1,1,1)_{12}$			
Parameter	Est.	S.E.		
$AR(1)$	-0.4379	0.1899		
$MA(1)$	0.5106	0.1805		
$SAR(1)$	-0.0440	0.0279		
$SMA(1)$	-0.9693	0.0101		
$SMA(2)$	-	-		

NB: The above models have been selected based on AIC and parsimony after fitting several competing models. S.E. and Est. respectively denote the standard error and estimates.

4.3.2 Model Diagnostics

In time series modelling, the selection of the best model fit to the data is directly related to whether the residual analysis is performed well. One of the assumptions of SARIMA (Seasonal ARIMA) model is that, for a good model, the residuals that fit the data well should be white noise, meaning that they should have mean zero, uncorrelated and portray unsystematic uniformly random variability over time (Chatfield, 1989). To validate the model for each station, the results of the diagnostics tests conducted are displayed in Table 4.5, Table 4.6 and Figure 4.5.

From Figure 4.5, the plotted standardised residuals versus time exhibits no pattern, revert around mean zero and variance one, for all the stations. The plotted ACF of residuals versus lag lies within the confidence interval, exhibiting stationarity. Finally, the p-values for the Ljung-Box test are above 0.05% for all lag orders, showing that there is no significant departure from white noise for the residuals. In summary, the plots indicate that the model fits the data or is suitable for the data, for all the stations of the Limpopo Province.

Table 4.5: Box-Ljung test results from SARIMA model for all the stations

Weather Station	**Box-Ljung test** Lag	Chi-Square	P-Value
Macuville	24	13.9590	0.9476
Mara	24	12.3850	0.9752
Marnits	24	235.7400	0.0000
Groendraai	24	9.5876	0.9960
Letaba	24	16.0840	0.8850
Pietersburg	24	8.8557	0.9979
Nebo	24	12.9430	0.9670

NB: *In each case, we fail to reject the null hypothesis of no serial correlation. Corresponding graphs for the residuals are displayed in Figure 4.5. The null hypothesis of normality is rejected.*

Table 4.6: Jarque-Bera test results from SARIMA model for all the stations

	Macuville	Mara	Marnits	Groendraai	Letaba	Pietersburg	Nebo
J-B	23947 (0.0000)	11118 (0.0000)	4656 (0.0000)	3581.5000 (0.0000)	28820 (0.0000)	3075 (0.0000)	224139 (0.0000)

Test statistic (chi-square) with the p-value in parenthesis for each station.

4.3.3 Forecasting SARIMA

According to Hendry and Clements (2003), a forecast is any statement about the future. Forecasting is one of the main objectives of this study. Using the constructed models for forecasting future rainfall, the model that fits the data well is bound to give good forecasts.

The performance of SARIMA models for all the rainfall stations is now evaluated by forecasting the data one step prediction for the year 2016 (Jan–Dec) to indicate the model's adequacy, performance and for comparison purposes. Using the selected models, the 12 months of forecast are shown in Table 5.1 and Figure 5.2, in the Appendices. It appears from Figure 5.2 that the se-

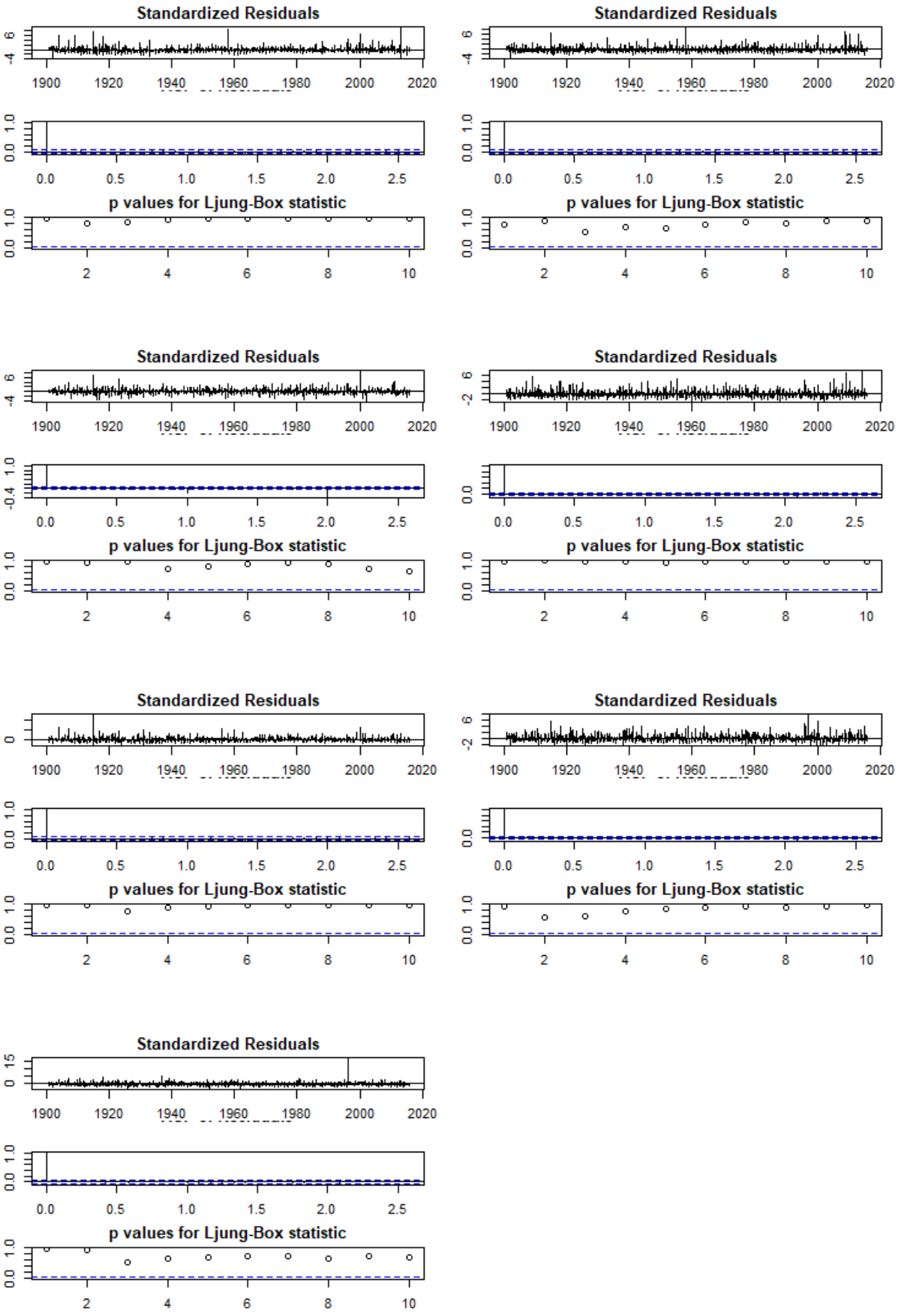


Figure 4.5: Time series diagnostics plots (model resduals) for all 7 stations of Limpopo Province in order.

lected models are very well suited for predicting the future development of the Limpopo rainfall, as the differences between the actual data and forecast data are very small. We, therefore, conclude that the SARIMA model is suitable for predicting the rainfall patterns in the 7 stations of the Limpopo province, using the data from January 1900 through to December 2016.

4.4 Presentation of ETS State Space

The exponential smoothing methods are useful for making forecasts and require no assumptions about the correlations between successive values of the time series, and to make prediction intervals for forecasts. The prediction interval requires that the forecast errors should be uncorrelated and normally distributed with mean zero and constant variance. The exponential smoothing state space models that adequately fit the data, were described as the ETS (A,N,A). Table 4.7 shows the parameters used in generating these models. These parameters were chosen to generate data that is reasonably realistic.

4.4.1 Estimation

Table 4.7: ETS(A,N,A) model parameters for all stations

Weather Stations	Macuville	Mara	Marnits	Groendraai	Letaba	Pietersburg	Nebo
α	0.0213	0.0220	0.0126	0.0004	0.0215	0.0024	0.0033
β	—-	—-	—-	—-	—-	—-	—-
γ	0.0001	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
σ	46.9194	42.4605	36.3133	40.6777	48.1523	36.5520	40.1116
l_0	55.4688	72.2723	43.5442	50.0182	45.5100	43.6888	53.7261
S_0	34.2582	43.2108	38.0571	52.8875	50.8177	45.0066	56.7167
S_{-1}	22.9558	31.7871	25.2469	33.9977	28.2051	39.0296	51.0770
S_{-2}	-9.6538	-4.4012	-5.3700	-6.5757	-8.6764	-3.2075	5.2139
S_{-3}	-25.2010	-28.7924	-25.4122	-35.7155	-29.2070	-26.5825	-34.7094
S_{-4}	-33.2873	-37.8006	-32.0214	-40.4955	-37.0072	-35.2906	-45.0159
S_{-5}	-32.6125	-35.3556	-33.3362	-44.7396	-36.8472	-36.8182	-47.6919
S_{-6}	—-	—-	—-	—-	—-	—-	—-
AIC	20805.95	20527.95	20092.56	20408.53	20878.16	20110.79	20369.51
RMSE	46.6828	42.2464	36.1302	40.4726	47.9095	36.3677	36.9094

The parameters of ETS models were estimated using maximum likelihood estimation method (see Table 4.7). Both the smoothing parameters α, β, γ and damping parameter σ are restricted. This is customary in order that the equations are presented in weighted average format and the damping parameter is

constraint further to lie between 0.80 and 0.98, in order to facilitate the model estimation (Hyndman and Athanasopoulos, 2014).

4.4.2 Model Diagnostics

Residual analysis forms one of the important steps to validate model adequacy. The plots of the residuals for each fitted model show that all the residuals seem to be white noise and form a random pattern around horizontal line zero. This implies that the fitted models are able to capture all the dynamics in the series. But there is an evidence of some cyclic effects as well as occurrence of outliers, as mentioned earlier in the time plot of the original data (Montgomery et al., 2008; Wei, 2006; Chatfield, 1995).

The cyclic effects point to the epileptic nature of the rainfall at certain period within the year, while the presence of outliers is an indication that irregular higher rainfall also takes place at some time periods of the year in the Limpopo Province.

Table 4.8: Box-Ljung test results from ETS model for all the stations

Weather Station	**Box-Ljung test**		
	Lag	Chi-Square	P-Value
Macuville	24	109.4500	0.0000
Mara	24	20.1570	0.6878
Marnits	24	38.5400	0.0305
Groendraai	24	40.8080	0.0175
Letaba	24	36.3530	0.0507
Pietersburg	24	18.2610	0.7902
Nebo	24	23.5880	0.4853

NB: *In each case, we fail to reject the null hypothesis of no serial correlation. The null hypothesis of normality is rejected.*

Table 4.9: Jarque-Bera test results from ETS model for all the stations

	Macuville	Mara	Marnits	Groendraai	Letaba	Pietersburg	Nebo
J-B	299.8600 (0.0000)	11023 (0.0000)	10755 (0.0000)	3373.8000 (0.0000)	41798 (0.0000)	3338 (0.0000)	221780 (0.0000)

Test statistic (chi-square) with the p-value in parenthesis for each station.

4.4.3 Forecasting ETS State Space

ETS point forecasts are equal to the medians of the forecast distributions. For models with only additive components, the forecast distributions are normal, implying that the medians and means are equal. For ETS models with multiplicative errors, or with multiplicative seasonality, the point forecasts will not be equal to the means of the forecast distributions.

A big advantage of the models is that prediction intervals can also be generated (something that was not possible before). The prediction intervals will differ between models with additive and multiplicative methods (Hyndman et al., 2008). The point forecast results together with their actual values are shown in Table 5.2 for ETS state space models for all the seven stations of the Limpopo Province.

4.5 Comparison of SARIMA and ETS State Space

ETS and ARIMA models provide complementary approaches to the problem of time series forecasting. While the former framework is based on a description of trend and seasonality in the data, the latter one aims to describe the autocorrelations in the data. There is the idea that ARIMA models are more general than exponential smoothing models. Actually, the two classes of models are complimentary, each with its strengths and weaknesses. While linear expo-

nential smoothing models are all special cases of ARIMA models, the non-linear exponential smoothing models have no equivalent ARIMA counterparts. There are also many ARIMA models which have no exponential smoothing counterparts. In particular, every ETS model is non-stationary while ARIMA models can be stationary (Ramos et al., 2015).

Montgomery et al. (2008) indicated that the accuracy measures are used to compare different methods of modelling time series. The smaller the value of the accuracy measure, the better the fit of the model.

The results of these criteria are presented in Table 4.10. The forecasting ability of the state space model is examined with respect to the best fitting SARIMA model for comparing the forecasting ability of the models. From the forecasts evaluation results presented in Table 4.10, the RMSE indicates that the state space local level model with deterministic seasonal outperforms the Box-Jenkins model in other stations. The forecasting results show that the SARIMA model outperforms the state space model in other stations.

Table 4.10: Comparing SARIMA$(p,d,q)(P,D,Q)_{12}$ and Exponential Smoothing State Space (ETS) models

Weather Station	**SARIMA**	RMSE	**ETS**	RMSE
Macuville	$SARIMA(1,0,1)(1,1,1)_{12}$	46.1821	$ETS(A,N,A)$	46.6828
Mara	$SARIMA(1,0,1)(1,1,1)_{12}$	42.2690	$ETS(A,N,A)$	36.1302
Marnits	$SARIMA(2,0,2)(1,1,0)_{12}$	43.6340	$ETS(A,N,A)$	42.2464
Groendraai	$SARIMA(2,0,1)(2,1,1)_{12}$	40.0161	$ETS(A,N,A)$	40.4726
Letaba	$SARIMA(1,0,2)(1,1,2)_{12}$	47.7713	$ETS(A,N,A)$	47.9095
Pietersburg	$SARIMA(1,0,0)(0,1,1)_{12}$	36.3008	$ETS(A,N,A)$	36.3677
Nebo	$SARIMA(1,0,1)(1,1,1)_{12}$	40.0261	$ETS(A,N,A)$	39.9093

NB: *The smaller the root mean square error (RMSE), the better the fitted model.*

4.6 Chapter Summary

Chapter 4 presented and analysed the results pertaining to rainfall in the Limpopo Province. The behaviour of the data, the stationarity tests and seasonality on the data were first presented, using the ADF and the HEGY test. Both methods where fitted to the data from different stations, they were estimated using maximum likelihood. In each station the best method was chosen based on the value of RMSE, the smaller the value, better the fitted model.

Chapter 5

Conclusion and Recommendations

5.1 Conclusion

Time series analysis is an important tool in modelling and forecasting rainfall. Two time series analysis techniques were applied in this study, the state space models based on exponential smoothing and the Box-Jenkins modelling method. Seasonal ARIMA models and ETS state space models were developed. Both models were found to be adequate for forecasting rainfall patterns in the Limpopo Province. The diagnostic checking confirms the adequacy of the models. The study compared ETS (Exponential Smoothing) Model and SARIMA (Seasonal Autoregressive Integrated Moving Average) Model with estimation of RMSE (Root Mean Square Error). Also included were the criteria AIC (Akaike's Information Criteria) values to check which model deploys the best results for forecasting rainfall in the seven stations of the Limpopo

Province.

The performance of the model is evaluated with reference to Root Mean Squared Error (RMSE) in different stations.

From Table 4.10, the comparative analyses revealed the following:

In Macuvile Station, the best models chosen were $SARIMA(1,0,1)(1,1,1)_{12}$ and $ETS(A,N,A)$, amongst others, based on the AIC. The $SARIMA$ model performed better than the ETS state space model, since the former accurately forecasted rainfall with less errors than the latter.

In Mara Station, the best models chosen were $SARIMA(1,0,1)(1,1,1)_{12}$ and $ETS(A,N,A)$, amongst others, based on the lowest value of the AIC. The ETS state space model performed better than the $SARIMA$ model, since the former accurately forecasted rainfall with less errors than the latter.

In Marnitz Station, the best models chosen were $SARIMA(2,0,2)(1,1,0)_{12}$ and $ETS(A,N,A)$, amongst others, based on the least value of the AIC. The ETS state space model performed better than the $SARIMA$ model, since the former accurately forecasted rainfall with less errors than the latter.

In Groendraai Station, the best models chosen were $SARIMA(2,0,1)(2,1,1)_{12}$ and $ETS(A,N,A)$, amongst others, based on the smallest values of AIC. The $SARIMA$ model performed better than the ETS state space model, since the former accurately forecasted rainfall with less errors than the latter.

In Letaba Station, the best models chosen were $SARIMA(1,0,2)(1,1,2)_{12}$ and $ETS(A,N,A)$, amongst others, based on the smallest values of AIC. The

$SARIMA$ model performed better than ETS state space model, since the former accurately forecasted rainfall with less errors than the latter.

In Pietersburg Station, the best models chosen were $SARIMA(1,0,0)(0,1,1)_{12}$ and $ETS(A,N,A)$, amongst others, based on least values of the AIC. The $SARIMA$ model performed better than ETS state space model, since the former accurately forecasted rainfall with less errors than the latter.

In Nebo Station, the best models chosen were $SARIMA(1,0,1)(1,11)_{12}$ and $ETS(A,N,A)$, amongst others, based on the least AIC values. The ETS state space model performed better than the $SARIMA$ model, since the former accurately forecasted rainfall with less errors than the latter.

The best models were used to forecast a 12-month rainfall data for the seven stations in the Limpopo Province. Both models, i.e. SARIMA and ETS state space, were found to be adequate for forecasting rainfall patterns. Diagnostic checking confirmed the adequacy of the models with respect to different stations of the Limpopo Province.

5.2 Evaluation of the Study

This study was set to address the problem identified in the first chapter. Therefore, this section revisits the objectives which were set to address the problem.

Objective 1: Determine the relationship between rainfall pattern in selected stations of the Limpopo province.
Most time series patterns can be described in terms of two basic classes of component: trend and seasonality. From the original plots (Figure 4.1) of all the rainfall stations, we can see that there is no trend, but seasonality exists. This

means that there is no relationship between the rainfall patterns in the various stations.

Objective 2: Propose a suitable model for forecasting the sporadic rainfall in all the selected stations of the Limpopo province using ETS state space and SARIMA models.
Several models were fitted from these two methods (ETS and SARIMA) in each station, and the best models were selected based on the AIC and parsimony.

Objective 3: Compare the forecast ability of ETS state space and SARIMA models.
In each rainfall station two models were selected from different methods. The same models were used to forecast rainfall. As shown in Table 4.1, the forecasting ability of the two models were compared using the RMSE, in order to choose the best model.

Thus, the study has addressed the three objectives as stipulated in section 1.3.3, i.e., (a) to establish the relationship between rainfall pattern in the seven selected stations of the Limpopo province; (b) to propose a suitable model for forecasting the sporadic rainfall in all these selected stations using the ETS state space model and SARIMA models; and (c) to compare the forecast ability of these two models.

5.3 Recommendations

The findings of this study could be used to forecast monthly rainfall for the upcoming years, for the selected seven stations of the Limpopo Province, analysed and compared in the dissertation. The results of the study provide information that could guide the decision makers when establishing strategies for proper

planning of agriculture, drainage system and other water resource applications in the Limpopo Province. The methods used in this study can be applied to other weather stations in South Africa or other countries, in predicting rainfall patterns in the future.

SARIMA and ETS state space models can be used by the Department of Agriculture and Rural Development in the Limpopo Province, to highlight the future rainfall patterns which could impact on the crop production to best suit the future situation. The rainfall forecasts can also be used by the Department of Water Affairs and Sanitation in Limpopo to come up with strategies and proper ways of water usage within the communities around the respective weather stations. These two methods can be used to predict extreme events such as floods and drought.

References

ABDUL-AZIZ, A., ANOKYE, M., KWAME, A., MUNYAKAZI, L., NSOWAH-NUAMAH, N., ET AL. (2013). Modeling and forecasting rainfall pattern in Ghana as a Seasonal ARIMA process: The case of Ashanti Region. *International Journal of Humanities and Social Science*, **3** (3), 224–233.

ABDULLAH, R., PRATIWI, E., ABDULLAH, L., DJA'WA, A., TENRIAWARU, A., AND FUJAJA, L. (2019). The role of economic in natural resources development in the city of Baubau. *In IOP Conference Series: Earth and Environmental Science*, volume 235. IOP Publishing, p. 012002.

AFRIFA-YAMOAH, E., SAEED, B., AND KARIM, A. (2016). SARIMA modelling and forecasting of monthly rainfall in the Brong Ahafo Region of Ghana. *World Environment*, **6** (1), 1–9.

AGRISA (2016). A raindrop in the drought. *Agri SA's status report on the current drought crisis*.

AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19** (6), 716–723.

AKAIKE, H. (1978). On the likelihood of a time series model. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **27** (3-4), 217–235.

AKINYEMI, O., AYENI, O., FAWEYA, O., AND IBRAHEEM, A. (2013). Statistical study of annual and monthly rainfall patterns in Ekiti State, Nigeria.

International Journal of Pure and Applied Sciences and Technology, **15** (2), 1.

AKPANTA, A., OKORIE, I., AND OKOYE, N. (2015). SARIMA modelling of the frequency of monthly rainfall in Umuahia, Abia State of Nigeria. *American Journal of Mathematics and Statistics*, **5** (2), 82–87.

AKRAM, M., HYNDMAN, R. J., ORD, J. K., ET AL. (2007). Non-linear Exponential Smoothing and positive data. Technical report, Monash University, Department of Econometrics and Business Statistics.

ALEXANDER, D. (1993). On the causes of landslides: Human activities, perception, and natural processes. *Environmental Geology and Water Sciences*, **20** (3), 165–179.

AMPAW, E. M., AKUFFO, B., LARBI, S. O., AND LARTEY, S. (2013). Time series modelling of rainfall in New Juaben Municipality of the Eastern Region of Ghana. *International Journal of Business and Social Science*, **4** (8).

ARIYARATHNA, P. AND ALIBUHTTO, M. (2016). Modeling monthly rainfall in Katunayake Region using Seasonal ARIMA model.

ASEMOTA, O. J. (2016). State Space versus SARIMA modeling of the Nigeria's crude oil export. *Sri Lankan Journal of Applied Statistics*, **17** (2).

BAGIROV, A. M., MAHMOOD, A., AND BARTON, A. (2017). Prediction of monthly rainfall in Victoria, Australia: Clusterwise Linear Regression approach. *Atmospheric Research*, **188**, 20–29.

BAKO, H. Y. (2014). *Forecasting pelagic fish in Malaysia using ETS State Space approach*. Ph.D. thesis, Universiti Tun Hussein Onn Malaysia.

BALDASSARRE, G., SCHUMANN, G., BATES, P. D., FREER, J. E., AND BEVEN, K. J. (2010). Flood-plain mapping: A critical discussion of deterministic and

probabilistic approaches. *Hydrological Sciences Journal–Journal des Sciences Hydrologiques*, **55** (3), 364–376.

BANERJEE, A., DOLADO, J. J., GALBRAITH, J. W., HENDRY, D., ET AL. (1993). Co-integration, error correction, and the econometric analysis of non-stationary data. *OUP Catalogue*.

BAUDOIN, M.-A., VOGEL, C., NORTJE, K., AND NAIK, M. (2017). Living with drought in South Africa: lessons learnt from the recent El Nino drought period. *International Journal of Disaster Risk Reduction*, **23**, 128–137.

BEKELE, F. (1997). Ethiopian use of ENSO information in its Seasonal forecasts. *Internet Journal of African Studies*, (2).

BENHIN, J. K. (2006). Climate change and South African agriculture: Impacts and adaptation options. Technical report, CEEPA discussion paper.

BOX, G. E. AND JENKINS, G. (1976). Time series analysis: Forecasting and control. *Time Series Analysis: Forecasting and Control. Revised Edition. Oakland, California: Editorial Holden-Day*.

BOX, G. E. AND JENKINS, G. M. (1970). Time series analysis: Forecasting and control Holden-Day. *San Francisco*, 498.

BOX, G. E., JENKINS, G. M., REINSEL, G. C., AND LJUNG, G. M. (2008). Time series analysis: Forecasting and control. *Hoboken, NJ*.

BROCKWELL, P. AND DAVIS, R. (1991). Time series: Data analysis and theory.

BROCKWELL, P. J. AND DAVIS, R. A. (2016). *Introduction to time series and forecasting*. Springer.

BROCKWELL, P. J., DAVIS, R. A., AND FIENBERG, S. E. (1991). *Time series: Theory and methods*. Springer Science & Business Media.

BROWN, R. G. (1959). *Statistical forecasting for inventory control*. McGraw/Hill.

BURNHAM, K. P. AND ANDERSON, D. R. (1998). Practical use of the information-theoretic approach. *In Model Selection and Inference*. Springer, pp. 75–117.

CALOIERO, T., COSCARELLI, R., AND FERRARI, E. (2018). Application of the Innovative Trend analysis method for the trend analysis of rainfall Anomalies in Southern Italy. *Water Resources Management*, **32** (15), 4971–4983.

CHATFIELD, C. (1989). Time series analysis: An introduction.

CHATFIELD, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **158** (3), 419–444.

CHONGE, M., KENNEDY, N., OMUKOBA, M., LUCY, M., AND FRANKLINE, T. (2015). A time series model of rainfall pattern of Uasin Gishu County. *IOSR J Math*, **11** (5), 77–84.

CONCHO VALLEY COUNCIL OF GOVERNMENTS (2012). Hazard mitigation plan update.

COOK, G. D. AND HEERDEGEN, R. G. (2001). Spatial variation in the duration of the rainy season in Monsoonal Australia. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **21** (14), 1723–1732.

CREVITS, R. AND CROUX, C. (2017). Forecasting using Robust Exponential Smoothing with Damped Trend and Seasonal components. *KBI_1741*.

CRYER, J. D. AND CHAN, K.-S. (2008). Time series analysis with applications in R, second edition springer science+ business media.

DEPARTMENT OF ENVIRONMENTAL AFFAIRS (2018). The Thabametsi case: Case no 65662/16 Earthlife Africa Johannesburg Minister of Environmental Affairs.

DEPARTMENT OF ENVIRONMENTAL AFFAIRS (2019). Minister of Environmental Affairs highlights progress on the implementation of the integrated strategic management of rhinoceros.

DICKEY, D. A. AND FULLER, W. A. (1979). Distribution of the estimators for Autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74** (366a), 427–431.

DICKEY, D. A. AND FULLER, W. A. (1981). Likelihood ratio statistics for Autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, 1057–1072.

DISASTER RELIEF EMERGENCY FUND OPERATIONS (2011). DREF final report South Africa: Floods, accessed: 15 April 2018.
URL: *https://reliefweb.int/sites/reliefweb.int/files/resources/MDRZA005dfr.pdf*

DISASTER RELIEF EMERGENCY FUND OPERATIONS (2014). DREF operations update South Africa: Limpopo floods, accessed: 15 April 2018.
URL: *https://reliefweb.int/sites/reliefweb.int/files/resources/MDRZA00601.pdf*

DOSIO, A. (2017). Projection of temperature and heat waves for Africa with an ensemble of CORDEX Regional climate models. *Climate Dynamics*, **49** (1-2), 493–519.

DRAGO, C. AND MASSA, E. (2019). Measuring and forecasting financial advisory demand using a Hybrid ETS-ANN model.

DU PLESSIS, J. AND BURGER, G. (2015). Investigation into increasing short-duration rainfall intensities in South Africa. *Water SA*, **41** (3), 416–424.

EL-SHAFIE, A., NOURELDIN, A., TAHA, M., HUSSAIN, A., MUKHLISIN, M., AND SOLOMATINE, D. (2012). Dynamic versus Static Neural Network model for rainfall forecasting at Klang River Basin, Malaysia. *Hydrology & Earth System Sciences*, **16** (4).

ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.

ETUK, E. H. AND MOHAMED, T. M. (2014). Time series analysis of monthly rainfall data for the Gadaref rainfall station, Sudan, by SARIMA methods. *International Journal of Scientific Research in Knowledge*, **2** (7), 320.

FAO (2004). Drought impact mitigation and prevention in the Limpopo River Basin. *Land and Water Discussion Paper 4*.

FLOODLIST (2017). Accessed: 13 February 2017.
URL: *http://floodlist.com/africa/nasa-measures-rainfall-ex-tropical-cyclone-dineo*

FLOODLIST (2019). Accessed: 14 April 2019.
URL: *http://floodlist.com/africa/cyclone-idai-eu-aid-mozambique-zimbabwe-malawi*

FLUGEL, W.-A., MARKER, M., MORETTI, S., RODOLFI, G., AND SIDROCHUK, A. (2003). Integrating geographical information systems, remote sensing, ground truthing and modelling approaches for regional erosion classification of semi-arid catchments in South Africa. *Hydrological Processes*, **17** (5), 929–942.

FRANSES, P. H. (1991). Seasonality, non-stationarity and the forecasting of monthly time series. *International Journal of Forecasting*, **7** (2), 199–208.

GARDNER, E. S. AND ACAR, Y. (2019). Fitting the Damped Trend method of Exponential Smoothing. *Journal of the Operational Research Society*, **70** (6), 926–930.

GARDNER, E. S., JR (1985). Exponential Smoothing: The state of the art. *Journal of Forecasting*, **4** (1), 1–28.

GARDNER, E. S., JR AND MCKENZIE, E. (1985). Forecasting trends in time series. *Management Science*, **31** (10), 1237–1246.

GIZAW, M. S. AND GAN, T. Y. (2017). Impact of climate change and El Niño episodes on droughts in Sub-Saharan Africa. *Climate Dynamics*, **49** (1-2), 665–682.

HAMILTON, J. D. (1994). *Time series analysis*, volume 2. Princeton New Jersey.

HARVEY, A. C. (1990). *Forecasting, Structural time series models and the Kalman Filter*. Cambridge University Press.

HENDRY, D. F. AND CLEMENTS, M. P. (2003). Economic forecasting: Some lessons from recent research. *Economic Modelling*, **20** (2), 301–329.

HIGAREDA, C. E. R., GUEVARA, L. I. P., LEÓN, S. C., ZAZUETA, J. G. S., MARTÍNEZ, J. C., ET AL. (2014). Analysis of rainfall variables trends and potential vegetation responses in Sinaloa, México. *American Journal of Climate Change*, **3** (03), 338.

HOLT, C. (1957). Forecasting seasonals and trends by Exponentially Weighted Moving Averages, ONR Memorandum (vol. 52), Pittsburgh, PA: Carnegie Institute of Technology. *Available from the Engineering Library, University of Texas at Austin*, **52**.

HTIKE, K. K. AND KHALIFA, O. O. (2010). Rainfall forecasting models using focused time-delay Neural Networks. *In International Conference on Computer and Communication Engineering (ICCCE'10)*. IEEE, pp. 1–6.

HU, W., TONG, S., MENGERSEN, K., AND CONNELL, D. (2007). Weather variability and the incidence of cryptosporidiosis: Comparison of time series Poisson Regression and SARIMA models. *Annals of Epidemiology*, **17** (9), 679–688.

HU, Z. (2008). Time series forecasting model for Chinese future marketing price of copper and aluminum.

HURVICH, C. M. AND TSAI, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, **76** (2), 297–307.

HYLLEBERG, S. (1995). Tests for seasonal unit roots general to specific or specific to general? *Journal of Econometrics*, **69** (1), 5–25.

HYLLEBERG, S., ENGLE, R. F., GRANGER, C. W., AND YOO, B. S. (1990). Seasonal Integration and Cointegration. *Journal of Econometrics*, **44** (1-2), 215–238.

HYNDMAN, R., KOEHLER, A. B., ORD, J. K., AND SNYDER, R. D. (2008). *Forecasting with Exponential Smoothing: The State Space approach*. Springer Science & Business Media.

HYNDMAN, R. J. AND ATHANASOPOULOS, G. (2014). Optimally reconciling forecasts in a hierarchy. *Foresight: The International Journal of Applied Forecasting*, (35), 42–48.

HYNDMAN, R. J. AND ATHANASOPOULOS, G. (2017). Forecasting: Principles and practice. otexts; 2014. *The book is freely available as an online book at www. OTexts. org/fpp. Alternatively, a print version is available*.

HYNDMAN, R. J. AND ATHANASOPOULOS, G. (2018). *Forecasting: Principles and practice*. OTexts.

HYNDMAN, R. J., KOEHLER, A. B., SNYDER, R. D., AND GROSE, S. (2002). A State Space framework for automatic forecasting using Exponential Smoothing methods. *International Journal of Forecasting*, **18** (3), 439–454.

IBRAHIM, L. AND DAUDA, U. (2016). Modelling monthly rainfall time series using ETS State Space and SARIMA models. *International Journal of Current Research*, **1** (1), 11–14.

IFFAT, A. (2009). A Neural Network approach to time series forecasting. *In Proceedings of the World Congress on Engineering, Vol. II, WCE*.

IWOK, I. A. AND DOOGA, M. (2017). Mixed Seasonal and Subset Fourier model with Seasonal Harmonics.

JAIN, G. AND MALLICK, B. (2017). A study of time series models ARIMA and ETS. *Available at SSRN 2898968*.

JAISWAL, R., LOHANI, A., AND TIWARI, H. (2015). Statistical analysis for change detection and trend assessment in climatological parameters. *Environmental Processes*, **2** (4), 729–749.

JARQUE, C. M. AND BERA, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of Regression residuals. *Economics Letters*, **6** (3), 255–259.

JOFIPASI, C. A. ET AL. (2018). Selection for the best ETS (error, trend, seasonal) model to forecast weather in the Aceh Besar District. *In IOP Conference Series: Materials Science and Engineering*, volume 352. IOP Publishing, p. 012055.

JURY, W. A. AND VAUX, H. (2005). The role of science in solving the world's emerging water problems. *Proceedings of the National Academy of Sciences*, **102** (44), 15715–15720.

KANTZ, H. AND SCHREIBER, T. (2004). *Nonlinear time series analysis*, volume 7. Cambridge University Press.

KAUR, S. AND RAKSHIT, M. (2019). Seasonal and Periodic Autoregressive time series models used for forecasting rainfall data.

KENDALL, S. AND ORD, K. (1990). Time series. *Edward Arnold*.

KIBUNJA, H. W., KIHORO, J. M., ORWA, G. O., AND YODAN, W. (2014). Forecasting precipitation using SARIMA model: A case study of Mt. Kenya Region. *Mathematical Theory and Modelling*, **4**, 50–59.

KOKILAVANI, S., PANGAYARSELVI, R., RAMANATHAN, S., DHEEBAKARAN, G., SATHYAMOORTHY, N., MARAGATHAM, N., AND GOWTHAM, R. (2020). Sarima modelling and forecasting of monthly rainfall patterns for coimbatore, tamil nadu, india. *Current Journal of Applied Science and Technology*, 69–76.

KOTIR, J. H. (2010). Climate change and variability in Sub-Saharan Africa: A review of current and future trends and impacts on agriculture and food security. *Environment, Development and Sustainability*, **13** (3), 587–605.

KRUGER, A. C. AND NXUMALO, M. (2017). Historical rainfall trends in South Africa: 1921–2015. *Water SA*, **43** (2), 285–297.

KRUGER, A. C. AND SHONGWE, S. (2004). Temperature trends in South Africa: 1960–2003. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **24** (15), 1929–1945.

KWIATKOWSKI, D., PHILLIPS, P. C., SCHMIDT, P., SHIN, Y., ET AL. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, **54** (1-3), 159–178.

LAVANYA, S., RADHA, M., AND ARULANANDU, U. (2018). Statistical Distribution of Seasonal rainfall data for rainfall pattern in TNAU1 Station Coimbatore, Tamil Nadu, India. *Int. J. Curr. Microbiol. App. Sci*, **7** (4), 3053–3062.

LE ROUX, J. (2011). Monitoring soil erosion in South Africa at a regional scale. *Agricultural Research Council-Institute for Soil, Climate and Water, Pretoria*.

LE ROUX, J., MORGENTHAL, T., MALHERBE, J., PRETORIUS, D., AND SUMNER, P. (2008). Water erosion prediction at a national scale for South Africa. *Water SA*, **34** (3), 305–314.

LIMPOPO ENVIRONMENTAL OUTLOOK REPORT (2016). Chapter 11: Planning for the future.

LINDESAY, J. AND JURY, M. (1991). Atmospheric circulation controls and characteristics of a flood event in Central South Africa. *International Journal of Climatology*, **11** (6), 609–627.

LJUNG, G. M. AND BOX, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, **65** (2), 297–303.

LONGOBARDI, A. AND VILLANI, P. (2010). Trend analysis of annual and seasonal rainfall time series in the Mediterranean Area. *International Journal of Climatology*, **30** (10), 1538–1546.

MAHSIN, M. (2012). Modeling rainfall in Dhaka Division of Bangladesh using time series analysis. *Journal of Mathematical Modelling and Application*, **1** (5), 67–73.

MALHERBE, J., ENGELBRECHT, F. A., LANDMAN, W. A., AND ENGELBRECHT, C. J. (2012). Tropical systems from the Southwest Indian Ocean making

landfall over the Limpopo River Basin, Southern Africa: A historical perspective. *International Journal of Climatology*, **32** (7), 1018–1032.

MAPONYA, P. AND MPANDELI, S. (2012). Impact of drought on food scarcity in Limpopo Province, South Africa. *African Journal of Agricultural Research*, **7** (37), 5270–5277.

MAPONYA, P. AND MPANDELI, S. (2016). Drought and food scarcity in Limpopo Province, South Africa. *In 2nd world irrigation forum*. pp. 6–8.

MARERA, D.-H. S.-V. (2016). *An application of Exponential Smoothing methods to weather related data*. Ph.D. thesis.

MASON, S. AND JURY, M. (1997). Climatic variability and change over Southern Africa: A reflection on underlying processes. *Progress in Physical Geography*, **21** (1), 23–50.

MAZVIMAVI, D. (2010). Investigating changes over time of annual rainfall in Zimbabwe. *Hydrology and Earth System Sciences*, **14** (12), 2671.

MONTGOMERY, J. M., LEE, T.-W., AND GRAY, S. K. (2008). Theory and modeling of light interactions with metallic nanostructures. *Journal of Physics: Condensed Matter*, **20** (32), 323201.

MURAT, M., MALINOWSKA, I., HOFFMANN, H., AND BARANOWSKI, P. (2016). Statistical modelling of agrometeorological time series by Exponential Smoothing. *International Agrophysics*, **30** (1), 57–65.

NATIONAL WEATHER SERVICE (2019). Accessed: 04 July 2019.
URL: *https://www.weather.gov/mrx/flood and flash*

NEWEY, W. K. AND WEST, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 777–787.

NIANG, I., RUPPEL, O., ABDRABO, M., ESSEL, A., LENNARD, C., PADGHAM, J., AND URQUHART, P. (2014). Africa climate change 2014: Impacts, adaptation, and vulnerability. part b: Regional aspects. contribution of working group ii to the fifth assessment report of the intergovernmental panel on climate change.

NICHOLLS, N. (1980). Long-range weather forecasting: Value, status, and prospects. *Reviews of Geophysics*, **18** (4), 771–788.

NICHOLSON, M., SMITH, A. C., STEWART, B., AND HOYE, R. (2018). *Sport management: Principles and applications*. Routledge.

NIEUWOLT, S., GHAZALLI, M. Z., AND GOPINATHAN, B. (1983). Agro-ecological Regions in Peninsular Malaysia.

NIRMALA, M. AND SUNDARAM, S. (2010). A Seasonal ARIMA model for forecasting monthly rainfall in Tamilnadu. *National Journal on Advances in Building Sciences and Mechanics*, **1** (2), 43–47.

ORD, J. K., KOEHLER, A. B., AND SNYDER, R. D. (1997). Estimation and prediction for a class of Dynamic Nonlinear Statistical models. *Journal of the American Statistical Association*, **92** (440), 1621–1629.

OSTASHCHUK, O. (2017). Time series data prediction and analysis. *Master's thesis. Czech Technical University in Prague Faculty of Electrical Engineering Department of Computer Science*.

OSTI, R., TANAKA, S., AND TOKIOKA, T. (2008). Flood hazard mapping in developing countries: Problems and prospects. *Disaster Prevention and Management: An International Journal*.

OTTO, F. E., WOLSKI, P., LEHNER, F., TEBALDI, C., VAN OLDENBORGH, G. J., HOGESTEEGER, S., SINGH, R., HOLDEN, P., FKAR, N. S., ODOULAMI,

R. C., ET AL. (2018). Anthropogenic influence on the drivers of the Western Cape drought 2015–2017. *Environmental Research Letters*, **13** (12), 124010.

PALMER, T. AND AINSLIE, A. (2002). Country pasture/forage resource profiles, grassland and pasture crops. FAO.

PANIGRAHI, S. AND BEHERA, H. S. (2017). A hybrid ETS–ANN model for time series forecasting. *Engineering Applications of Artificial Intelligence*, **66**, 49–59.

PANKRATZ, A. (1983). Forecasting with Univariate Box-Jenkins models: Concepts and cases. John Wily and Sons. *Inc. USA*.

PAPACHARALAMPOUS, G., TYRALIS, H., AND KOUTSOYIANNIS, D. (2018). Predictability of monthly temperature and precipitation using Automatic time series forecasting methods. *Acta Geophysica*, **66** (4), 807–831.

PAZVAKAWAMBWA, G. T. (2017). A time series forecasting model for Windhoek rainfall, Namibia.

PEGELS, C. C. (1969). Exponential forecasting: Some new variations. *Management Science*, 311–315.

PERRON, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, **54**, 1361–1401.

PETRIS, G., PETRONE, S., AND CAMPAGNOLI, P. (2009). Dynamic linear models. *In Dynamic Linear Models with R*. Springer, pp. 31–84.

PHAKULA, S. ET AL. (2016). *Modelling Intra-Seasonal rainfall characteristics over South Africa*. Ph.D. thesis, University of Pretoria.

PHILIPPON, N., ROUAULT, M., RICHARD, Y., AND FAVRE, A. (2011). The influence of ENSO on winter rainfall in South Africa. *International Journal of Climatology*, **32** (15), 2333–2347.

RAMOS, P. AND OLIVEIRA, J. (2016). A procedure for identification of appropriate State Space and ARIMA models based on time series cross-validation. *Algorithms*, **9** (4), 76.

RAMOS, P., SANTOS, N., AND REBELO, R. (2015). Performance of State Space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing*, **34**, 151–163.

SAKAMOTO, Y., ISHIGURO, M., AND KITAGAWA, G. (1986). Akaike Information Criterion statistics. *Dordrecht, The Netherlands: D. Reidel*, **81**.

SALHI, A., MARTIN-VIDE, J., BENHAMROUCHE, A., BENABDELOUAHAB, S., HIMI, M., BENABDELOUAHAB, T., AND PONSATI, A. C. (2019). Rainfall distribution and trends of the daily precipitation concentration index in Northern Morocco: A need for an adaptive environmental policy. *SN Applied Sciences*, **1** (3), 277.

SAYEMUZZAMAN, M. AND JHA, M. K. (2014). Seasonal and Annual precipitation time series trend analysis in North Carolina, United States. *Atmospheric Research*, **137**, 183–194.

SCHULZE, R. (2003). Design rainfall and flood estimation in South Africa.

SCHWERT, G. (2009). Quantitative micro software, eviews 7 user's guide ii. *Quantitative Micro Software, LLC, Irvine, California*.

SHONGWE, M. E., VAN OLDENBORGH, G., VAN DEN HURK, B., DE BOER, B., COELHO, C., AND VAN AALST, M. (2009). Projected changes in mean and extreme precipitation in Africa under global warming. part i: Southern Africa. *Journal of Climate*, **22** (13), 3819–3837.

SOUTH AFRICA (2019). Accessed: 16 May 2019.
URL: *http://www.geographia.com/south-africa/index.html*

SOUTH AFRICAN WEATHER SERVICE (2015). Annual report, accessed: 21 June 2018.
URL: *http://www.weathersa.co.za/Documents/AnnualReports/SAWS-Annual-Report-14-15.pdf*

STAGER, J. C., RYVES, D. B., KING, C., MADSON, J., HAZZARD, M., NEUMANN, F. H., AND MAUD, R. (2013). Late holocene precipitation variability in the summer rainfall Region of South Africa. *Quaternary Science Reviews*, **67**, 105–120.

STATISTICS SOUTH AFRICA (2007). South African Statistics, accessed: 28 June 2019.
URL: *http://www.statssa.gov.za/publications/SAStatistics/SAStatistics2007.pdf*

STOFFER, D. AND SHUMWAY, R. (2010). Time series analysis and its application (3rd edn, vol. 10, pp. 1441978658).

SUBBAIAH NAIDU, K. (2016). SARIMA modeling and forecasting of Seasonal rainfall patterns in India. *Int J Math Trends Technol (IJMTT)*, **38** (1), 15–22.

SWAIN, S., VERMA, M., AND VERMA, M. (2015). Statistical trend analysis of monthly rainfall for Raipur District, Chhattisgarh. *Int. J. Adv. Engg. Res. Studies/IV/II/Jan.-March*, **87**, 89.

TAKELE, R. (2012). *Statistical analysis of rainfall pattern in Dire Dawa, Eastern Ethiopia*. Ph.D. thesis, M.Sc. Thesis submitted to the Department of Statistics, Addis Ababa University.

TAKELE, R. AND GEBRETSIDIK, S. (2015). Prediction of long-term pattern and its extreme event frequency of rainfall in Dire Dawa Region, Eastern Ethiopia. *Journal of Climatology & Weather Forecasting*, 1–15.

TAYLOR, A. B. (2003). Ontogeny and function of the masticatory complex in go-

rilla: Functional, evolutionary, and taxonomic implications. *Gorilla Biology: A Multidisciplinary Perspective*, 132–189.

TODOKO, C. A. K. (2013). *Time series analysis of water consumption in the Hohoe Municipality of the Volta Region of Ghana*. Ph.D. thesis.

TSAY, R. S. (2005). *Analysis of financial time series*, volume 543. John Wiley & Sons.

UDDIN, K., CHAUDHARY, S., CHETTRI, N., KOTRU, R., MURTHY, M., CHAUDHARY, R. P., NING, W., SHRESTHA, S. M., AND GAUTAM, S. K. (2015). The changing land cover and fragmenting forest on the roof of the world: A case study in Nepal's Kailash Sacred Landscape. *Landscape and Urban Planning*, **141**, 1–10.

UNGANAI, L. AND KOGAN, F. (1998). Southern Africa's recent droughts from space. *Advances in Space Research*, **21** (3), 507–511.

UNITED NATIONS (2009). Millennium development goals report 2009 (includes the 2009 progress chart).

UNITED NATIONS OFFICE (2016). United nations office for the coordination of humanitarian affairs. *Connecting People and Data to Improve Lives*.

VALIPOUR, M. (2015). Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorological Applications*, **22** (3), 592–598.

VALIPOUR, M. (2016). Optimization of Neural Networks for precipitation analysis in a Humid Region to detect drought and wet year alarms. *Meteorological Applications*, **23** (1), 91–100.

VON STORCH, H. AND ZWIERS, F. W. (1988). Recurrence analysis of climate sensitivity experiments. *Journal of Climate*, **1** (2), 157–171.

WALKER, S. AND TSUBO, M. (2003). *Estimation of rainfall intensity for potential crop production on clay soil with in-field water harvesting practices in a Semi-arid Area*. Water Research Commission Pretoria.

WANG, S., FENG, J., AND LIU, G. (2013). Application of seasonal time series model in the precipitation forecast. *Mathematical and Computer Modelling*, **58** (3-4), 677–683.

WEESAKUL, U. AND LOWANICHCHAI, S. (2005). Rainfall forecast for agricultural water allocation planning in Thailand. *Science & Technology Asia*, 18–27.

WEI, W. W. (2006). Time series analysis. *In The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2*.

WINTERS, P. R. (1960). Forecasting sales by Exponentially Weighted Moving Averages. *Management Science*, **6** (3), 324–342.

WIREs CLIMATE CHANGE (2014). Wires Climate Change after 4 years: An editorial essay. *Wiley Interdisciplinary Reviews: Climate Change*, **5** (1), 1–5.

WOLDE-GEORGIS, T. (1997). El Nino and drought early warning in Ethiopia. *Internet Journal of African Studies*, (2).

YASMEEN, F. AND HAMEED, S. (2018). Forecasting of rainfall in Pakistan via Sliced Functional Times Series (sfts). *World Environment 2018*, **8** (1), 1–14.

YUSOF, F. AND KANE, I. (2012). Modelling monthly rainfall time series using ETS State Space and SARIMA models. *International Journal of Current Research*, **4** (9), 195–200.

YUSOF, F. AND KANE, I. L. (2013). Volatility modeling of rainfall time series. *Theoretical and Applied Climatology*, **113** (1-2), 247–258.

ZHOU, Y., ZHANG, Q., AND SINGH, V. P. (2014). Fractal-based evaluation of the effect of water reservoirs on hydrological processes: The dams in the Yangtze River as a case study. *Stochastic Environmental Research and Risk Assessment*, **28** (2), 263–279.

ZIERVOGEL, G., NEW, M., ARCHER VAN GARDEREN, E., MIDGLEY, G., TAYLOR, A., HAMANN, R., STUART-HILL, S., MYERS, J., AND WARBURTON, M. (2014). Climate change impacts and adaptation in South Africa. *Wiley Interdisciplinary Reviews: Climate Change*, **5** (5), 605–620.

Appendix

Table 5.1: Point forecast from SARIMA model for all the stations

Weather Stations	Macuville	Mara	Marnits	Groendraai	Letaba	Pietersburg	Nebo
Jan	107.5468 (9.3)	90.7612 (43.4)	52.7207 (65.1)	110.8547 (96.1)	76.8907 (24.8)	84.2572 (40.3)	100.1463 (55.8)
Feb	58.0160 (87)	62.4977 (29)	28.0992 (63.8)	90.4044 (31.9)	77.2656 (23.2)	65.5738 (58)	82.0173 (153.7)
Mar	37.8888 (133.3)	52.1204 (120.9)	79.3854 (58.9)	103.6173 (148.8)	50.4312 (99.2)	57.1451 (96.1)	64.6652 (71.3)
Apr	29.2844 (6.2)	29.6241 (6)	37.1720 (3)	38.5296 (12)	28.6776 (3)	31.2460 (9)	42.0122 (0)
May	5.0953 (3.1)	10.3393 (40.3)	-0.8120 (12.4)	12.1101 (21.7)	4.2774 (21.7)	10.4331 (24.8)	13.1722 (6.2)
Jun	4.0423 (3)	5.2080 (9)	1.0615 (9)	5.3591 (15)	-0.7115 (0)	5.2191 (3)	4.4760 (3)
Jul	2.5244 (0)	3.7312 (0)	-0.3176 (0)	6.4938 (15.5)	1.6732 (18.6)	3.0446 (0)	3.2866 (3.1)
Aug	0.2490 (0)	3.7570 (0)	-0.1916 (0)	5.6002 (0)	0.5687 (0)	4.2384 (0)	5.8651 (0)
Sep	8.2299 (0)	12.3121 (0)	16.8599 (0)	14.5160 (3)	8.9153 (18)	12.2590 (3)	15.0684 (3)
Oct	20.1205 (15.5)	34.4578 (15.5)	18.7061 (9.3)	44.1866 (58.9)	27.2328 (18.6)	35.5961 (27.9)	56.3891 (46.5)
Nov	53.7490 (54)	73.8370 (63)	53.8630 (66)	91.8633 (111)	64.9803 (84)	81.1684 (63)	100.1987 (192)
Dec	61.2472 (99.5)	80.9303 (189.1)	87.2733 (99.6)	104.3405 (127.1)	83.6673 (102.3)	82.2612 (133.3)	112.0181 (86.8)

NB: *Point forecast values with the Actual values in the parenthesis for each station.*

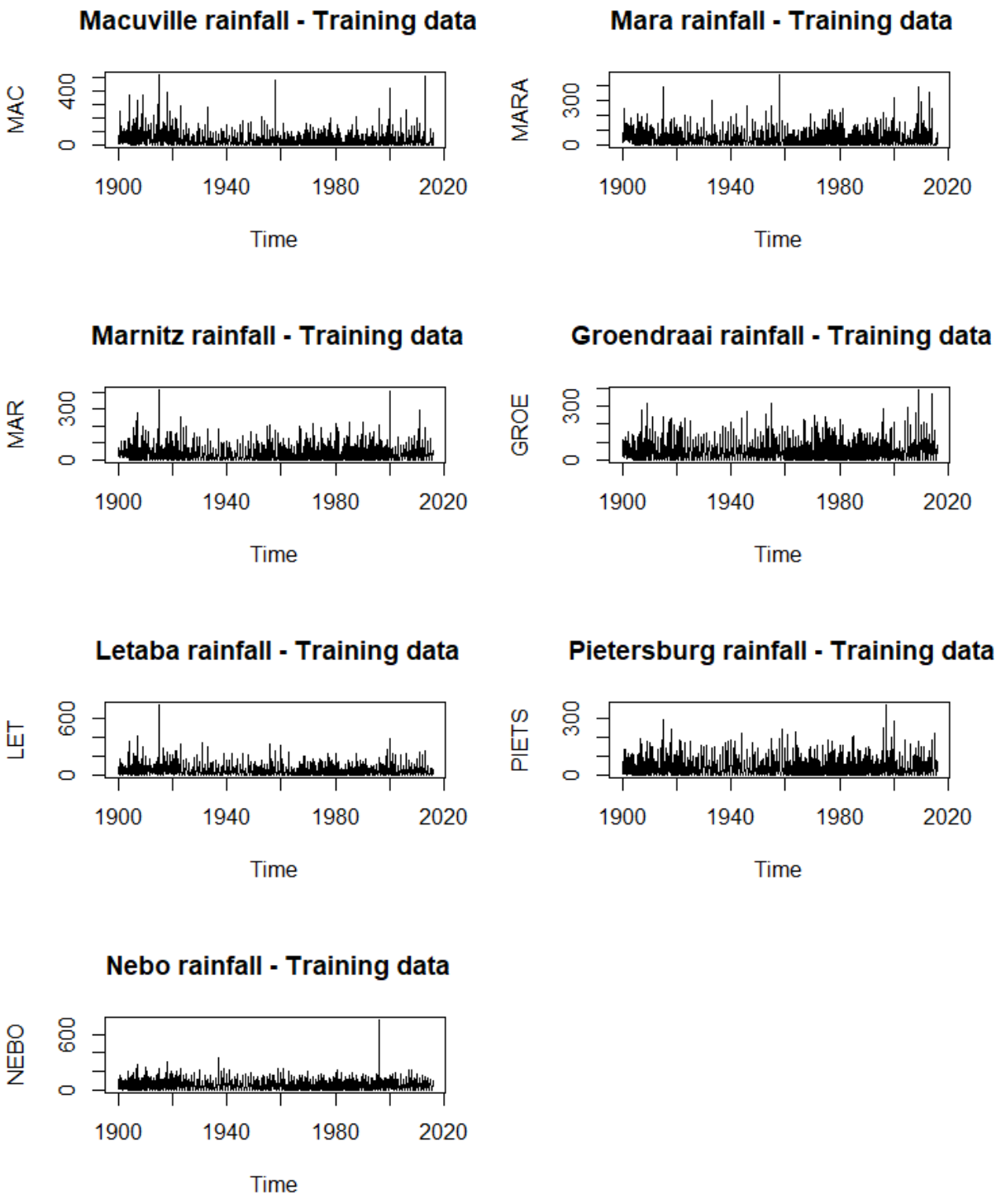


Figure 5.1: Training Time series plots for monthly rainfall data from Jan. 1900 to Dec. 2015 for all 7 stations of Limpopo Province.

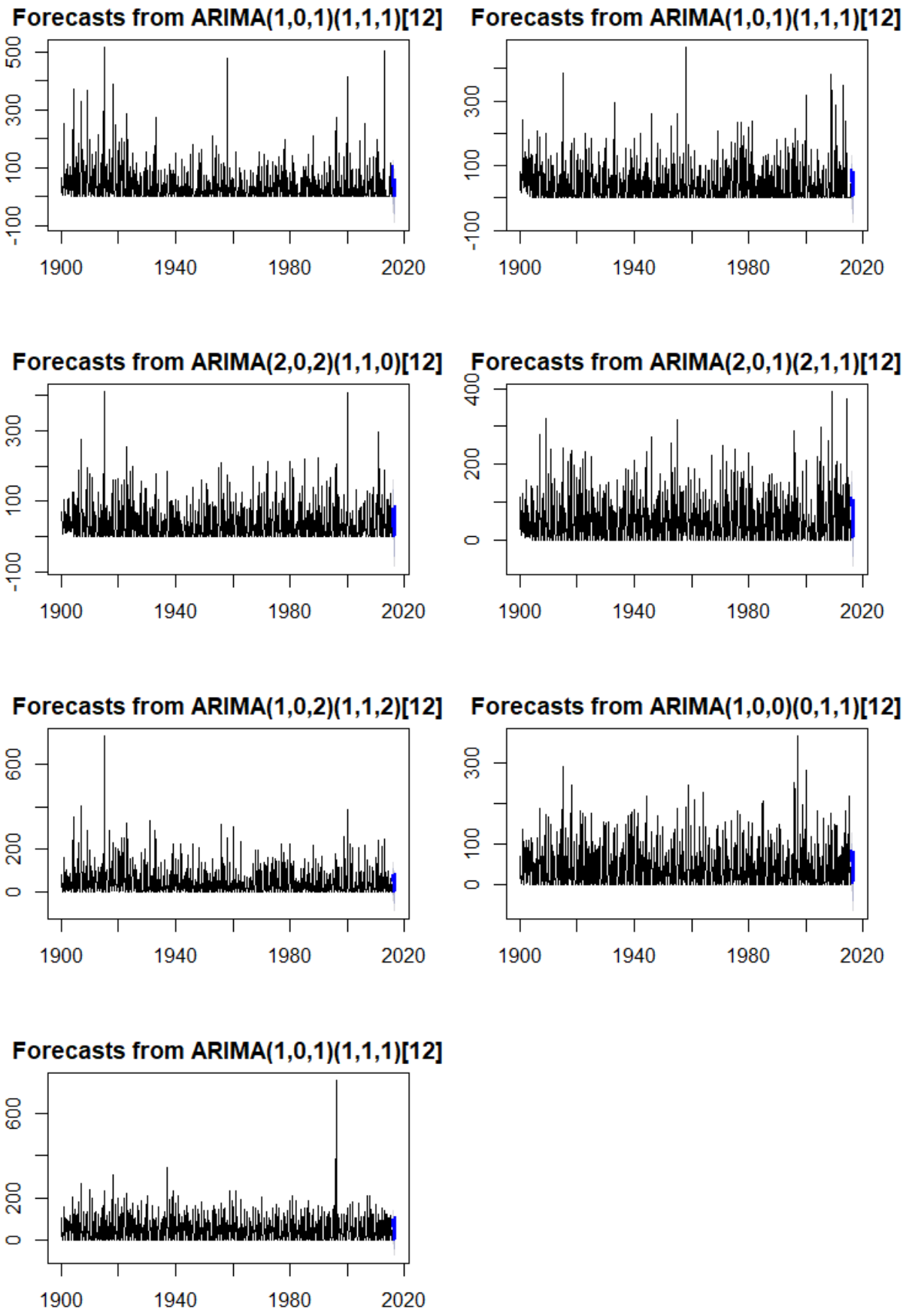


Figure 5.2: Point forecast from SARIMA model for all the stations

Table 5.2: Point forecast from ETS model for all the stations

Weather Stations	Macuville	Mara	Marnits	Groendraai	Letaba	Pietersburg	Nebo
Jan	82.7770 (9.3)	93.6577 (43.4)	90.5075 (65.1)	113.8100 (96.1)	86.8180 (24.8)	82.7572 (40.3)	108.9629 (55.8)
Feb	71.4477 (87)	68.0560 (29)	71.5996 (63.8)	95.8102 (31.9)	76.7454 (23.2)	67.6439 (58)	81.4052 (153.7)
Mar	45.2015 (133.3)	55.2715 (120.9)	55.5681 (58.9)	74.0826 (148.8)	53.5391 (99.2)	58.6935 (96.1)	68.5729 (71.3)
Apr	18.8502 (6.2)	28.1920 (6)	28.7617 (3)	37.3663 (12)	24.2949 (3)	29.8336 (9)	38.0039 (0)
May	1.7910 (3.1)	11.8032 (40.3)	12.8878 (12.4)	13.6827 (21.7)	1.3600 (21.7)	9.4846 (24.8)	12.1609 (6.2)
Jun	-1.9032 (3)	5.0160 (9)	6.9690 (9)	6.0469 (15)	-4.5229 (0)	5.3020 (3)	2.4958 (3)
Jul	-3.5060 (0)	3.0902 (0)	5.5801 (0)	5.2800 (15.5)	-2.5809 (18.6)	2.4914 (0)	1.8403 (3.1)
Aug	-4.1823 (0)	0.6737 (0)	6.8886 (0)	9.4883 (0)	-2.7535 (0)	4.0155 (0)	4.5122 (0)
Sep	3.9120 (0)	9.6610 (0)	13.4960 (0)	14.3057 (3)	5.0487 (18)	12.7162 (3)	14.8221 (3)
Oct	19.4454 (15.5)	34.0467 (15.5)	33.5314 (9.3)	43.4512 (58.9)	25.5850 (18.6)	36.0914 (27.9)	54.7389 (46.5)
Nov	52.0673 (54)	70.2490 (63)	64.1504 (66)	84.0241 (111)	62.4550 (84)	78.3540 (63)	100.5968 (192)
Dec	63.3730 (99.5)	81.6485 (189.1)	76.9630 (99.6)	102.9029 (127.1)	85.0722 (102.3)	84.2998 (133.3)	106.2458 (86.8)

NB: *Point forecast values with the Actual values in the parenthesis for each station.*