TIME SERIES MODELLING OF WATER EVAPORATION FROM SELECTED DAMS IN THE LIMPOPO PROVINCE OF SOUTH AFRICA

by

MMANYAKU GOITSEMANG PHASHA

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 Science)

at the

UNIVERSITY OF LIMPOPO

SUPERVISOR: Prof D Maposa

2022

Declaration

I, Mmanyaku Goitsemang Phasha, declare that the dissertation hereby submitted to the University of Limpopo for the degree of Master of Science in Statistics has not previously been submitted by me for a degree at this or any other university; that it is my work in design and in execution, and that all material contained herein has been duly acknowledged.

Phasha MG (Ms)	25-04-2022
Surname, Initials (title)	**Date**

Abstract

Water is a precious natural resource and one of the most vital substance for sustainability of life . The increase in water evaporation is a major problem where factors such as high temperature and minimum rainfall are the contributing factors. The aim of the study was to perform time series modelling of water evaporation from the selected dams in the Limpopo province South Africa. A daily evaporation time series data was used in the study with variables such as temperature and rainfall. Daily water evaporation rate time series data was differenced to make the data series stationary and Dickey-Fuller test was used to test the stationarity of the data series. The Autoregressive Conditional Heteroskasticity (ARCH) and Generalized Autoregressive Conditional Heteroskasticity (GARCH) model was performed on the water evaporation time series data from the selected dams. Vector Autoregression (VAR) was used to determine the relationship between the variables evaporation, rainfall and temperature. Identification of time series models was done using the autoregressive integrated moving average (ARIMA). The best ARIMA models were selected based on the autocorrelation function (ACF) and partial autocorrelation function (PACF), and the smallest value of Bayseian Information (BIC). The best models selected for each dam are: Mokolo dam, ARIMA (1, 1, 2) model; Ga-Rantho dam, ARIMA (1, 1, 2) model; Leeukraal DeHoop dam, ARIMA (1, 1, 1) model and Luphephe dam, ARIMA (2, 1, 3) model. The correlation coefficient, coefficient of determinant (R^2) and root mean square (RMSE) were used to determine the performance of the model. The water evaporation time series data from the selected dams was forecasted using the best selected ARIMA models from the selected dams and then predicted for the next 3 years, where the results showed a positive constant water evaporation rate.

Dedication

I would like to dedicate the research project to my son Bohlokwa Ditheto Phasha and my mother Tubake Constance Phasha for giving me the motivation, support and courage when I was going through challenges with my research project. I would like to also dedicate this work to my family and friends whom motivated me to focus and work hard on this research project. I genuinely appreciate them for being there for me.

Acknowledgments

I would like to firstly thank the Almighty God for giving me the strength and knowledge throughout the research project and being able to complete this dissertation.

I would like to take this time to show my sincere gratitude and appreciation to my supervisor Prof D Maposa for his light, support, guidance and encouragement throughout the dissertation. I would like to also thank my best friend Trevor Makhobelele for the support towards this dissertation, and Alphuse Rachidi for borrowing me his laptop to use for the research project. I would also like to extent my gratitude to South African Weather Service (SAWS) for playing a huge role in this dissertation in terms of providing data. Special thanks to my family and friends.

Contents

List of Tables

List of Figures

List of Abbreviation and Acronomys

ACF - Autocorrelation Function

ADF - Augmented Dickey-Fuller

AIC - Akaike's Information Criterion

ANN - Artificial Neural Network

AR - Autoregressive

ARMA - Autoregressive Moving Average

ARCH - Autoregressive Conditional Heteroscedastisity

ARIMA - Autoregressive Integrated Moving Average

AWS - Automatic Weather Station

BIC - Bayesian Information Criterion

ET - Evaporation

GAMs - Generalised Additive Models

GARCH - Generalised Autoregressive Cnditional Heteroscedasticity

GDP - Gross Domestic Product

GIS - Geographic Information System

GRNN - Generalised Regression Neural Network

HADA - Highaswan Dam Authority

LAS - Large Aperture Scintillometer

MA - Moving Average

MAE - Mean Absolute Error

MAD - Mean Absolute Deviation

MAPE - Mean Absolute Percent Error

MFE - Mean Forecast Error

MGARH - Multivariate Generalised Autoregressive Conditional Heteroscedastis

MLP - Multi- layer Perception

MLP - Multiple Layer Regression

MODIS - Moderate resolution imgaging spectorradiometer system

MSE - Mean Square Error

NSE - Nash Sutcliffe Cofficient of Efficiency

PACF Partial Autocorrelation Function

PML - Penman-Montheith-Launing

PMP - Penma- Monteith- Plamer

RBF-NN - Radial Basis Function Nneural Network

RMSE - Root Mean Square Error

RS - Remote Sensing

SA - South Africa

SARMA - Seasonal Autoregressive Moving Average

SARIMA - Seasonal Autoregressive Integrated Moving Average

SEBS - Surface energy balance system

SPEI - Standardised Precipition Evaporation Index

SRV - Support Vector Regression

VAR - Vector Autoregressive

WI - Willmott's Index

WMA - Weighted Moving Average

WRF - Weather Research and Forecasting

WRSI - Water Requirement Satisfaction Index

Chapter 1

Introduction and background

1.1 Introduction

The quickly growing population of the world is putting a gradual pressure on fresh water supplies. This is in addition to the pressure placed on water supply due to the rainfall timing and climate change. The amount of rainfall, climate change and other elements such as temperature and wind speed have a threat to water supply and profitable agricultural production, meaning that improvement in the management of water resources and efficient water use are essential (Chami and Moujabber, 2016; Sule and Ajala, 2017). The need to provide a growing population with enough fresh water is increasing water scarcity and declining water quality. This has brought sustainability of water resources management into consideration of the global development agenda (Mekkonen and Hoekstra, 2010).

Water quality is a worldwide problem that affects the lives of human beings fundamentally (Taheri et al., 2014). Water evaporation is a major contribution towards water scarcity. Evaporation plays an important role in water resources planning, operation and management because a lot of water is lost through evaporation especially in large reservoirs. Higher evaporation creates more arid environment, while low trend of evaporation result in more humid environment (Sule and Ajala, 2017). Since evaporation is a challenging factor to the accessibility of quality water, its modelling using time series analysis is needed for water resources management, irrigation scheduling, agricultural

management and reservoir operation. There are several methods for modelling hydrological variables such as computing techniques, physical-based modelling and stochastic time series modelling (Dabral et al., 2017). Time series is a pattern of data over time and there is an equal interval between all data, whether is daily, weekly, monthly or yearly data. Time series analysis is used for decision making in many hydrological processes and operation systems. It aims to model the stochastic mechanism of hydrologic trends, as well as to forecast the future values of the trends (Taheri et al., 2014).

1.2 Background

Worldwide, freshwater is an essential resource for all people and majority of them rely on groundwater for water supply and irrigated agricultural production. Globally energy production and industries are the largest water consumers accounting for 80-90% of water consumption along with evaporation loss from reservoirs. With a growing population and climate change, pollution and insufficient water recharge lead to declining groundwater levels (Salem et al., 2019; Yue et al., 2018). South Africa is a country situated at the southern part of the African continent, covering about 1.2 million square kilometers. South Africa consists of nine provinces namely: Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, North-West, Northern Cape, Western Cape. South Africa with a population rate of 59.31 million and a 1.28 yearly change in 2020 shows that there is an increase in the population growth rate, hence the demand for water supply is expected to increase.

According to Chami and Moujabber (2016); Hensley et al. (2019), South Africa is ranked as the 30^{th} driest and a water-scarcity country in the world, with unpredictable rainfall and a diverse range of climate change affecting water infiltration rate and water security. Frequency of drought occurrence is dominant in the semi-arid climate. South Africa has 22 water source areas situated in five provinces (Eastern Cape, KwaZulu Natala, Limpopo, Mpumalanga, Western Cape), with mean annual rainfall of 490 mm which is only half of the global average rainfall (Diamond and Jack, 2018). The country's water source areas are the sources of most major river streams and dams. Agriculture in South Africa is a major source of income, and majority of the population depends on agricultural activities. The water scarcity, high level of temperature and low level of rainfall in parts of the country, is a disaster threat to the peoples needs and the economy in general (Hensley et al., 2019).

Limpopo province is a province located in the northern part of South Africa and is made up of five districts and 25 municipalities. Limpopo province is a semi-arid region with low unreliable rainfall. The low rainfall has negative impact on the agricultural sector, water accessibility and industrial production (LDARD, 2015). The climate in the Limpopo province ranges from high level of rainfall to high level of dryness and climate in the province is often characterized by unavoidable change in extreme weather events (Musetha, 2016). According to Musetha (2016) and Machete (2019) climate change in the Limpopo province lead to raising temperatures, reducing the occurrence of rain and its timing. With raising temperatures, there is also a high probability of evaporation during that period.

Several rural communities in Limpopo province depend on agricultural production and water supply as basic needs. However, adverse climate conditions have a bearing on their needs for accessibility to food supply and water resource(Musetha, 2016). Agricultural production and water access remain as the source of the livelihood for most of the rural communities. Industries and agricultural production are sources of employment to more than 60% of the people and also contributing about 30% of gross domestic product (GDP). Therefore, low water accessibility due to high evaporation rate may affect these sectors leading to high unemployment rate in agricultural production and industries which might result in high level of poverty and health problems (Maponya, 2012; Nhemachena and Hassan, 2007).

Sekhukhune Ga-Rantho, Lephalale Mokolo, Sekhukhune Leeukraal DeHoop and Thohoyandou Luphephe are some of the dams located in the different districts of the Limpopo province. These dams are to be studied in the present study. Dams play an important role in the rural communities by providing water resource to the people, mining companies and agricultural production. Since water is a precious natural resource which is often affected by high temperature, humidity, wind speed, unreliable rainfall and evaporation, modelling water evaporation is important towards better planning and management of the water resource (Diamond and Jack, 2018; Hughes, 2019; Mosase et al., 2019).

1.3 Problem statement

Water is a precious natural resource and one of the most vital substances for sustainability in life. In many parts of the world, the shortage of water from

dams due to evaporation existed in the past and is still existing in some countries such as Botswana, Mozambique, Namibia, South Africa and Zimbabwe. Therefore, this problem must be managed to forestall its future adverse effects. Water resources are a major source of economic development for most African countries (Naabil et al., 2017). The availability of fresh water is becoming an increasingly significant problem worldwide. This problem is much more serious in arid regions that are faced with a severe shortage of fresh water (El-Ghonemy, 2012; Gorjian S. and Ghobadian, 2015). In the recent years, the water shortage problem has gained much attention due to climate change and global warming Baydaronglu and Kocak (2014). Climate change is one of the major factors attributed to the occurrence of water evaporation. According to Behrouzil and Chini (2017) and Brutsaert (2013) evaporation is the water loss from a liquid surface of a water body to the atmosphere. Evaporation increases with high wind speed, high temperature, pressure and low humidity. Water losses by evaporation in reservoirs reduce the efficient use of water in agriculture by methods such as well-built water pipes or drip irrigation and have an influence on the economy (Martinez-Granadose et al., 2011).

According to Wine et al. (2019) and Martinez-Granadose et al. (2011), in recent years agricultural, industrial and urban water demands have drastically increased in many countries, leading to shortage and fierce competition for water resources. In arid and semi-arid climates, loss of water due to evaporation causes low agricultural production that can lead to financial stress for farmers mostly in times of droughts (Gleick, 2014).

South Africa is one of the driest countries in the world due to extreme heat caused by temperature rise, and with a low rainfall sequence that makes managing the water resources a difficult task (Botai et al., 2018; Hensley et al., 2019; Meissner et al., 2018). Uneven rainfall distribution pattern over the country is an additional problem that worsens the access to availability of water resources. As the population and the economy grow, the necessity for fresh water will also increase (Kuun, 2009; Motoshita et al., 2018).

In most parts of the Limpopo province evapotranspiration exceeds rainfall and annual average temperature. According to Koppen-Geiger climate classification system, Limpopo province is determined to be the hot semi-arid region Mzezewa et al. (2010). Limpopo is a very dry region and rainfall is highest in the high-lying areas in the south than in the north (Masupha and Moeletsi, 2018; Shabalala et al., 2019). The short rainfall season and high hot temperature in the province decreases the level of water access and lowers

the agricultural standards (Mosase and Ahiablame , 2018; Shabalala et al., 2019).

1.4 Motivation

Globally, water resources are experiencing pressure due to rising demand from several social and economic driving forces. Water is an important source for human survival and ecosystem health. In arid and semi-arid areas like Limpopo River basin in South Africa, water is a precious product (Maponya, 2012). The limited accessibility of water has a negative impact towards the livelihoods of people and animals in farms and Kruger National Parks in the Limpopo province. Majority of people in the province depend on ground water, water from rivers and dams. With high temperature and uneven rainfall this makes it difficult for people to engage in agricultural production where most people depend on agricultural production as their source of income and for food consumption. The access of water for basic domestic use is also affected and industries that use water are also affected in the areas, which increases the rate of unemployment in the province (Mapholi, 2018; Ziervogel et al., 2006). It is very important that water security is well maintained in the province so that water supply can be more efficient. Mostly in previous studies, water evaporation was estimated directly from water streams, rivers and dams using traditional methods, hence modelling water evaporation from dams using time series analysis will help in planning, securing and managing the scarcity of water access by finding the best models to model and predict the water evaporation from the dams (Hughes, 2019; Mosase et al., 2019).

1.4.1 Aim

The aim of the study is to perform time series modelling of water evaporation from selected dams in the Limpopo province of South Africa.

1.4.2 Objectives

The objectives of the study are to:

i. Perform an autoregressive conditional heteroscedastcity (ARCH) and generalised autoregressive conditional heteroscedasticity (GARCH) times series modelling of water evaporation from the selected dams.

ii. Apply the vector autoregression (VAR) model to determine the relationship between evaporation and the explanatory variables rainfall and temperature.

iii. Apply hybrid linear (ARMAX) and multivariate GARCH (MGARCH) in modelling water evaporation.

iv. Perform a comparative analysis of the various time series models used in the study.

v. Forecast water evaporation in the Limpopo province from selected dams using the best selected time series models.

1.5 Significance of the study

Modelling water evaporation loss from dams in the Limpopo province will improve awareness in the water users leading to improved water security in the province and increased agricultural production in the farming industry. The time series techniques used in the data analysis will add to the body of knowledge in the use of statistical techniques in water evaporation data.

1.6 Structure of the dissertation

This section presents the structure of the dissertation. The dissertation comprises five chapters as well as the reference list and appendices. The rest of the dissertation is organised as follows: Chapter 1 presents the problem statement, motivation, aim and objectives of the study. Chapter 2 presents literature review. The literature review covers previous studies related to time series modelling of water evaporation from other countries and in South Africa. Chapter 3 gives the research methodology of the study. In this chapter detailed approach on how the data was obtained and methods used to analyse the data are outlined. Chapter 4 presents the data analysis, results and discussion. All the interpretations and discussions of the results are made in this chapter. In Chapter 5, the conclusion, limitations, delimitation, recommendations and areas for future research directions are provided.

Chapter 2

Literature review

2.1 Introduction

This chapter reviews previous studies on modelling time series of water evaporation, factors contributing to evaporation such as rainfall, temperature, humidity and wind speed. The effects of climate change towards water accessibility are also reviewed. The chapter also presents related methods used in time series with applications to water evaporation in South Africa, Africa and other countries worldwide.

2.2 Evaporation worldwide

A study was conducted by Althoff et al. (2019) with the aim to evaluate six different methods for estimating evaporation in order to select the most suitable method to use in hydrological models for water balance in reservoirs in the state of Ceará. The methods tested were Penman, Priestley-Taylor Deruim-Keijman, Kohler-Nordenson-Fox, Brutsaert-stricker and de-Bruim. The methods presented good performance when tested for water balance during the dry seasons. The Priestley-Taylor was found to be the most accurate method, since the data from simulated water balance with evaporation estimated by this method were the closest to the balance data observed from measures of reservoir level.

Globally, evaporation loss from reservoirs are estimated to be greater than

the combined consumption from industrial and domestic water uses. By fusing remote sensing and modelling approach, Zhao and Gao (2019) conducted a study to develop a novel method to accurately estimate the evaporation losses from 721 reservoirs in the contiguous United States. Penman equation was used to model the evaporation rate where the lake storage was considered. Validation results using in-situ observation suggest that this approach can significantly improve the accuracy of the simulated monthly reservoir evaporation rate. The results showed that the long term averaged annual evaporation volume from these 721 reservoirs is equivalent to the annual public water supply. An increasing trend of the evaporation rate and a slightly decreasing trend of total surface area were both detected during the study period. The total evaporation showed an insignificant trend, with significant spatial heterogeneity.

Water is the most important substance for sustainability of life on earth. The maximum amount of water loss from reservoirs occurs through evaporation, hence it is very important to know the dynamical system that governs the evaporation process. Baydaronglu and Kocak (2014) conducted a study using trajectory method to obtain a differential equation from reconstructed phase space of evaporation time series. The trajectory method was a success after it was applied to obtain the dynamical system that represents the periodic pattern of evaporation process.

Evaporation of water from reservoirs, rivers and agricultural fields results in major losses of critical water resources, especial in arid regions of the world. Dawood et al. (2013) did an investigation to reduce evaporation losses from water reservoirs. Trash of polyethylene with different densities was used as floating cover to the water filling cylindrical container with 8 cm diameter and led to reduction in the evaporation rate. With the method used, reducing the trash density it will reduce the evaporation rate, using trash density of 800 kg/m^3 will reduce evaporation rate by 57%. In a separate research, Craig and Hancock (2004) conducted a study to evaluate the effectiveness of chemical monolayers floating covers and shade structure in reducing dam evaporation. Evaporation was assessed using high precision pressure sensor transducer to measure small changes in the dam height. The evaporation rate was calculated as the residual in the dam water balance, taking into consideration in-flaw and out-flow, and seepage which is assumed to be the same as the night-time loss. In the warm semi-arid environments, night-time evaporation is less than the daytime evaporation rate.This method proved a successful and roust standard method for assessing the evaporation rate of Australia farm dams.

2.3 Evaporation in Africa

The impact of climatic conditions in predicting evaporation was explored from a reservoir. Allawi et al. (2019) conducted a study aimed at investigating the ability of radial basis function neural network (RBF-NN) and support vector regression (SRV) methods to develop an evaporation rate prediction model for a tropical area at the Laying Reservoir, Johor River, Malaysia.

Egypt is one of the countries that is experiencing limitation of water resources. Hassan (2018) investigated the use of remote sensing (RS) and geographic information system (GIS) techniques to calculate monthly evaporation rates. Surface energy balance system (SEBS) method using terra moderate resolution imaging spectroradiometer (MODIS) satellite earth observation data was used in the study. Atmospheric parameters predicted from weather research and forecasting (WRF) model were also used to estimate monthly evaporation rate using harbeck equation which is used by high aswan dam authority (HADA). The estimated evaporation rate by HADA is slightly higher than SEBS and MOD16ET. A high correlation was found between SEBS and MODI16ET which showed a good indication that only one method can be used to estimate the monthly evaporation rate.

Lake Victoria in state country is the largest fresh water lack in Africa. The water level of Lake Victoria is determined by its water balance, consisting of precipitation on the lake, evaporation from the lake, inflow from tributary rivers and lake outflow controlled by two hydropower dams. Vanderkelen et al. (2018) contacted a study to present a water balanced model for Lack Victoria using state of the art remote sensing observation and high-resolution reanalysis downscaling. Precipitation is the main cause of seasonal and interannual Lake Victoria level fluctuations. The results showed that the 2004-2005 drop in lack level can be about half attributed to a drought in the Lack Victoria Basin and about half to an enhanced outflow, highlighting the sensitivity of the lack level to human operations at the outflow dam.

In the Lake Nasser, state country evaporation is considered an important factor of the water balance system that causes a huge loss of lake's waters. Hassan (2013) estimated evaporation rate for Lake Nasser using the surface balance approach on remote sensing technology. With the method, the evaporation rate estimated during satellite overpass over the lake is instantaneous. Evaporation friction method was also used to estimate the daily rate from the instantaneous evaporation rate. The surface energy balance combined with remote sensing data showed a promising evaporation rate es-

timation for large water bodies, that could lead to more accurate monitoring of evaporation rate in the lake area.

2.4 Evaporation in South Africa

Accurate estimation of evaporation losses is important to manage the river resources efficiently. McKenzie and Craig (2001) conducted a study to initiate the improved estimation of river losses downstream of Vanderkloof Dam in South Africa and to develop a methodology for estimating evaporation losses from South African rivers. Theoretical losses were estimated from a measured evaporation rate, multiplied by the water surface area and an appropriate riparian vegetation area. Hydraulic modelling was used to perform a dynamic water balance to verify the initial loss estimate. Evaporation rate from a flowing river were found to be in the same order of magnitude as a pan evaporation data. The variation in evaporation losses was due to the changes in surface area with flow.

Gwate (2018) conducted a study comparing the performance of the Penman-Monteith-Launing (PML) and Penman-Monteith-Palmer (PMP) evapotranspiration (ET) models, over mesic grasslands in two study sites. Routine meteorological data from scientific-grade automatic weather station (AWS) was used. The two models were validated using ET derived from large aperture scintillometer (LAS). The PML model, performed well at both sites with root mean square error (RMSE) within 20% of the mean daily observed ET. The PML model was better to simulate observed ET compared to PMP model. Model prediction in the grassland could be improved by integrating the soil evaporation component in the PMP model, while PML model could be improved by careful choice of the number of days to be used in the determination of the fraction soil evaporation.

Droughts and global warming have raised major concerns for the agricultural sector, especially to farmers who rely on rain-fed farming. Masupha and Moeletsi (2018) conducted a study calculating the standardised precipitation evaporation index (SPEI) and water requirement satisfaction index (WRSI) to assess drought on a 120-day maturing maize crop. The results showed that 40-54% SPEI of the agricultural seasons, a mild drought condition was experienced. However, WRSI results indicated that stations in the driest regions of the catchment experienced mild drought corresponding with satisfactory crop performance every season. The results further showed an overall mild moderate drought in the near-future with SPEI decreasing

below -1.5. In the far-future the conditions are expected to change, whereby the crop performance predicted significantly in drier conditions ($p<0.05$).

Jovanovic et al. (2011) conducted a study comparing evaporation from endemic vegetation-Renosterveld and a dryland wheat/fallow cropping system. The study was carried out in the mid-reaches of the Berg River catchment, South Africa. Total evaporation was measured to be higher in Renosterveld than in wheat during the rainy winter season and in the dry summer season total evaporation from Renosterveld was limited soil water supply and vegetation was under water stress. Spatial variability of total evaporation from both wheat/fallow land and Renosterveld was estimated using surface energy balance algorithm for land (SEBAL) model. Scintillometer measurements were used to determine basal crop coefficients for long term (20 years) simulation with HYDRUS-1D model to assess temporal variability in total evaporation.

2.5 Evaporation and time series worldwide

Machekposhti et al. (2018) conducted a study that was to simulate and model the climatic variable, evaporation, using stochastic methods. Based on autogressive integrated moving average (ARIMA) model, the auto-correlation and partial auto-correlation methods, assessment of the parameters and type of models, the appropriate models to forecast evaporation were obtained. The results of the 10 years mean annual evaporation forecasts showed an increase in evaporation rate. In a separate study, Qasem et al. (2019) modelled monthly evaporation data using wavelet support vector regression (SVR) and wavelet artificial neural networks (ANN) in Tabriz (Iran) and Antalya (Turkey) stations. The results from the study showed that ANN provided reasonable trends for evaporation modelling at both Tabriz and Antalya stations.

Sun et al. (2015) conducted a study about time series models that are useful in estimating and forecasting of reference evapotranspiration series and their changes. The time series models considered were the generalised autoregressive conditional heteroscedasticity (GARCH) and seasonal autoregressive moving average (SARMA). The Akaike's information criterion (AIC) was used in selecting the final SARMA model and the results showed that it was efficient for modelling the monthly mean total daily evaporation series. According to Engle test, heteroscedasticity was present in the residuals of the SARMA model. Hence GARCH model was used for modelling reference

evaporation series. The GARCH model showed the ability to remove the heteroscedasticity from the reference evaporation residuals.

Fathian (2019) conducted a study aiming to demonstrate the ability of non-linear time series approaches to provide adequate modelling of streamflow. In the study, first two and three-regime (SETAR) models were used to model the mean behaviour of daily streamflow. The results showed that the hybrid SETRA- GARCH models performed better than the models without GARCH component. The results also showed that the use of non-linear SETAR and GARCH improves streamflow modelling efficiency by capturing the heteroscedasticity in the residuals of nonlinear threshold time series. In a separate study, Fathian (2019) used a hybrid of linear (ARMAX) variables and nonlinear GARCH, as well as the multivariate GARCH (MGARCH) time series models to model water level time varying and variance. The fitted models identified streamflow, temperature, precipitation, wind speed, relative humidity and day length as the factors affecting water level's time varying mean and variance.

Dabral et al. (2017) conducted a study aiming to estimate reference evapospiration using the Penman-Moetheith FAO-56 method. Cube root transformation was applied to smooth the data and stabilise the variance in the monthly ET_0 data. The collected data from year 1961-2005 was used for time series modelling, and data from 2001-2005 was used for model validation. Turning point and Mann-kendall test were used at 5% significance level for identifying trend component. The time series data was made stationary by removing the periodic component, and autoregressive (AR), moving average (MA), autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) were investigated for modelling a dependent stochastic component. ARIMA (12,1,1) model was found to be the best fit model based on the minimum value Bayesian information criterion (BIC) statistic and correlation coefficient and Nash-Suttclift coefficient indicated high degree of model fitness. Portmanteau test and Box-cox transformation were applied to series, (a_t) of independent stochastic component for independence and normalisation checking. The monthly ET_0 was validated with a time series of 5 years and forecasted for the years 2006 to 2050.

A study was conducted by Tezel and Buyukyildiz (2016) aiming to use Artificial neural network method (ANNs), multi-layer perception (MLP), radial basis function network (RBFN) and ϵ-support vector regression (SVR) to estimate monthly pan evaporation. Temperature, relative humidity, wind speed and precipitation data from 1972 to 2005 were used as input variables,

while pan evaporation data was used as an output. Romanenko and Meyer methods were considered for comparison. The algorithm performance was assessed via mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination, R^2. The results showed that ANN algorithms and ϵ-SVR had similar results and both methods were found to perform better than the Romanenko and Meyer methods.

Quantum-behaved particle swarm optimisation algorithm embedded into a multi-layer perception technique was developed to estimate evaporation rates over a daily forecast horizon. Evaporation data from years 2012-2014 for Thlesh meteorological station in Northern Iran was used. The predictive accuracy of the MLP-QPSO model was evaluated with existing hybrid MLP-PSO and standalone MLP model. The results were evaluated with respect to statistical performance criterion: the mean absolute error (MAE), root mean square error (RMSR), Willmott's index and Nash-Sutcliffe coefficient. Taylor diagrams were used to assess the level agreement between the forecast and observed evaporation data. The results showed that the hybrid MPL-QPSL model was an optimal forecasting tool applied for estimating daily pan evaporation than the MLP-PSO and the standalone model (Ali et al., 2018).

2.6 Evaporation and time series in Africa

Sule and Ajala (2017) conducted a study to use ARIMA models to forecast pan-evaporation data from Osogbo, southwest Nigeria. Regression analysis was done on the data and the autocorrelation indicated non-stationarity on the data. The AIC, BIC, as well as diagnostics of residuals confirmed that ARIMA (3, 4, 3) was a good fit for both short-term data forecast and data generated for pan-evaporation. Evaporation series was estimated from 2013 till 2062, and the results showed that with increasing evaporation trend the reservoir will not be able to serve the various benefiting towns after the year 2038.

Water level forecasting is important for the water catchment management specifically for flood warning systems. Arbain and Wibowo (2012) conducted a study in malaysia to predict water level with input variables monthly rainfall and rate of evaporation taken from Dungun River, Terengganu Malaysia. ARIMA and ANN models were used to predict the water level. Since the rainfall data contained imperfect characteristic data, the pre-processing data was made to the original data. After some experiment the results showed that the ANN with cleansing rainfall data gives better performance than

ARIMA and ANN models without cleansing the data.

Issaka (2015) used the vector autoregression (VAR) model to examine the dynamic relationship between rainfall and temperature time series data in Kassena-Nankana Municipality. A univariate ARCH-LM test and Ljung-Box test showed that the model is free from conditional heteroscedasticity and serial correlation. The study concluded that there is a bi-directional relationship between rainfall and temperature, showing that rainfall is helpful in explaining a considerable amount of the forecast uncertainty in temperature and vice versa.

Sebbar et al. (2020) conducted a study aiming to estimate daily evaporation from EL AGRAM Dam reservoir Jijel, East of Algeria using generalized regression neural network (GRNN) model. Four measured climatic variables data for a period of 13 years from 2003 to 2015 were used. For developing the models, four input variables measured at daily namely: daily maximum air temperature, daily minimum air temperature, daily wind speed and daily relative humidity were used. The performance of the models were analysed using the RMSE, MAE, Willmott's Index (WI), and correlation coefficient (R) statistics. The GRNN model was compared to multiple linear regression (MLR) with respect to their capability for modelling daily evaporation. The results obtained showed evaporation could successfully be estimated using the GRNN model. The GRNN model which used all the four input variables was the best model among all other tested models.

2.7 Evaporation and time series in South Africa

Makungo and Odiyo (2017) tested the ability of coupled linear and non-linear system identification model in estimating groundwater level. Daily groundwater levels for four boreholes, rainfall and evaporation data covering the period 2005-2014 were used. Correlation coefficient R, coefficient of determination R^2, RMSE, percent bias (PBIAS), Nash Sutcliffe coefficient of Efficiency (NSE) and graphical fits were used to evaluate the model performance. The results indicated that the model was able to estimate groundwater levels.

Forecasting extreme hydrological events is critical for drought risk and efficient water resource management in semi-arid environment. Mathivha et al. (2020) conducted a study aimed to forecast drought conditions in semi-arid region. The standardised precipitation evaporation index (SPEI) was used as a drought-quantifying parameter. Forecasting of the SPEI was achieved

by using generalized additive model (GAMs) at 1, 6, and 12-month time scale. To reduce time series complexities, time series composition was done and variable selection was done using Lasso. Four models where developed to forecast drought namely: GAM, Ensemble empirical mode decomposition (EEMD)-GAM, ARIMA-GAM and forecast quantile regression average (fQRA). The results showed that the first two time scales FQRA, forecasted the data better than the other models, while GMAs were the best at 12-month time scale. Root mean square error values of 0,0599, 0.2609 and 0.1809 were presented by FQRA and GAM at the 1,6, 12-month time scales. The results demonstrated the strength of GAM in short- term and medium-term drought forecasting.

Chapter 3

Research Methodology

3.1 Introduction

In this study time series models are used, and this chapter aims to give a detailed description of time series methods and its models. The description for testing stationarity and unit root are described. Test for auto-correlation and heteroscedasticity are described. Time series models and methods such as ARIMA, ARCH, GARCH and VAR models are to be defined. Description of Box-Jenkins and residual analysis, model selection and forecasting are to be discussed.

3.2 Data source and study area

The study will utilise secondary data for analysis of water evaporation from selected dams in the Limpopo province, obtained from Agro Climatology (AC). The data recorded is a daily data and consist of variables such as maximum temperature, minimum temperature, rainfall and evaporation. The selected dams and data record period are as follows: Lephalale: Mokolo dam (2008-2018), Sekhukhune: Leeukraal DeHoop dam (2008-2018), Sekhukune: Ga-Rantho (2008-2018) and Thohoyandou: Luphephe dam (2008-2018).

3.3 Time series

Time series is a set of observations arranged chronologically, a sequence of observations usually ordered in time. The time series sequence of data points or observations $\{X_1, X_2, X_3, ..., X_n\}$ is being recorded at a specific time, where the random variable X_1 represent the value taken by the series at first time point, variable X_2 represent the value series at second time point and so on. A common notation specifying a time series X that is indexed by natural numbers is written as:

X_t=$X_1, X_2, ..., X_n$.

Time series is denoted by X=$\{X_t$, t ϵT$\}$ where T is the index and is refereed as a stochastic process.

If T is continuous, then there is a continuous time series.

If T is discrete then there is a discrete time series and T= R that vary over a set of integers. The time series is written as ...,$x-_1, x_0, x_1, x_2, ...$ (Adhikari and Agrawal, 2013; Shumway et al., 2000).

The purpose of using time series is to find the occurring patterns which might be advantageous in estimating future values of the time series. Time series it is applied in many areas such as in the financial and economy sectors.

3.4 Time series analysis

Time series analysis is a pattern of well-defined data points measured at consistent time interval period, it examines the changing data often with the objective of predicting the future occurrences. Time series analysis is to describe the history of movements in time of some variable at a particular site. Time series analysis uses statistical methods to analyse time series data and extract meaningful statistics and characteristics about the data (Velicer and Velicer, 2003).

3.5 Time series plot

A time series plot is a graphical illustration of time series data, where on the x-axis the time-increments or date are plotted and on the y-axis the corresponding value that are going to be measured are plotted. Time series plots

are very useful in illustrating how the value of the thing that are interested to be measured changes over time. On the time series plot patterns of time series can be identified, (Adhikari and Agrawal, 2013; Montgomery et al., 2015) .

3.6 Time series components

The trend is the compound of a time series that shows variation of low occurrence in a time series, the high and medium frequency fluctuation have been sifted. A time series is affected by four main components which can be separated from the observed data. The four components of time series are as follows: Trend variation (T), Seasonal variation (S), Cyclical variation (C) and Irregular (I) (Adhikari and Agrawal, 2013).

3.6.1 Trend variation

Trend variation is a long term movement in a time series data, it is a time series that increases, decreases or stagnate over a long period of time. A trend can be positive or negative reliant on whether the time series shows an increasing long-term pattern or a decreasing long-term pattern. It shows irregular effects and is a reflection of the underlying level, it present influences such as population growth, price inflation and general economic changes (Adhikari and Agrawal, 2013).

3.6.2 Seasonal variation

Seasonal variation in a time series in which the data experiences consistent and predictable fluctuations within a year during the season. The important factors causing seasonal variation are climate, weather conditions, customs and traditional habits etc (Adhikari and Agrawal, 2013).

3.6.3 Cyclical variation

The cyclical variation in a time series describes the medium-term changes in the series, caused by outcomes which repeat in cycle, it is a pattern that exist when data display a rise and a fall that are not of a stable interval. The period of a cycle extend over a longer period of time, normally two or more years. Most of the economic and financial time series show some cyclical variation (Adhikari and Agrawal, 2013).

3.6.4 Irregular

Irregular component is what remains after the seasonal and trend components of a time series have been predicted and removed.The fluctuation in a time series that exist after considering the consistent effects random variations in data or due to unforeseen events such as strikes, earthquake, floods and revolution (Adhikari and Agrawal, 2013).

3.7 Time series decomposition

Time seris decomposition is presented and illustrated in a constructive result, is used in forecasting business and economic data. Decomposition is useful in analysis of an observed time series through inference about underlying, latent components that may have physical interpretations. The aim of time series analysis is to isolate the influence of each of the four components of the actual series. The multiplicative time series model is used to analyse the influence of these four components. The multipicative model is based on the idea that the actual values of time series X can be found by multiplying all four components. The multiplicative time series is defined as:

$$X_t = T_t \times S_t \times C_t \times I_t. \tag{3.1}$$

Another model that can be used is the additive model given by:

$$X_t = T_t + S_t + C_t + I_t. \tag{3.2}$$

Where X_t is the observation and X_t, S_t, C_t and I_t are respectively the trend, seasonal, cyclical and irregular variation at time t, (Adhikari and Agrawal, 2013; Hipel and McLeod, 1994).

3.8 Stationary and non-stationary time series

3.8.1 Stationary time series

A stationary time series is the procedure where statistical properties such as mean, variance and autocorrelation are all constant and does not depend on time, and if strictly periodic variations have been removed. For the time series data to be stationary, the observed plot should be a horizontal straight line, (Adhikari and Agrawal, 2013; Montgomery et al., 2015). There are two

types of stationary process namely: strictly stationary and weakly stationary time series.

Time series has a constant mean defined as:

$$\mu_{Y_t} = E(X_t) = \int_{-\infty}^{\infty} xf(x)dx \tag{3.3}$$

and a constant variance defined as:

$$\sigma_t^2 = Var(X_t) = \int_{-\infty}^{\infty} (x - \mu_{x_t})^2 f(x)dx. \tag{3.4}$$

Strictly stationary time series

A process x(t), t=0,1,2, ... is said to be strictly stationary if the joint distribution of $x_{t-s}, x_{t-s+1}, ..., x_t..., x_{t+s-1}, x_{t+s}$ is independent of t for all s. For a strictly stationary process the joint distribution of any possible set of random variables from the process is independent of time (Adhikari and Agrawal, 2013; Cochrane, 2005; Hipel and McLeod, 1994).

Weak stationary time series

A stochastic process is said to be weakly stationary of order k if the statistical moments of the process up to that order depend only on time difference and not upon the time of occurrences of the data being used to estimate the moments. Weakly stationary x(t),t=0,1,2,... is second order stationary if it has time independent mean and variance and the corvariance values $\text{Cov}(x_t, x_t - s)$ depends only on s, (Adhikari and Agrawal, 2013; Shumway et al., 2000).

White noise

A generated series collected of uncorrelated random variables, w_t, with mean 0 and finite variance σ_w^2. White noise is a time series generated from uncorrelated variables is used as a model for noise in engineering applications, this process is denoted as $w_t \sim wn(0, \sigma_w^2)$ (Shumway et al., 2000).

3.8.2 Non-stationary time series

A time series is said to be non-stationary when the time series depends on time, when time changes also the time series change. When a time series data is non-stationary, firstly the data should be transformed into stationary data so that further statistical analysis can be completed on the de-trend stationary data. There are different transformation methods that can be used to make a time series data stationary. The transformation methods such as: Differencing, Log transformation, Square root transformation, Arcsine transformation and Power transformation (Ihaka, 2005; Manuca and Savit, 1996; Zhang, 2016).

The reason for making transformation is to stabilise the variance, to make the seasonal effect additive and to make the data normally distributed. The transformation method to be used is differencing, it stabilises the mean and it gives reliable results.

3.8.3 Stationary through differencing

Differencing is a method of transforming a time series dataset, it can be used to remove the series dependence on time. This method is used when a time series model is not stationary so that it can be stationary (Brownlee, 2017; Reinert, 2010). Differencing a time series data is denoted by:

$$\nabla^k = \nabla(\nabla^{k-1}X)_t. \tag{3.5}$$

The operator ∇ represent difference operation. The procedure can be applied several times until a time series is stationary, the first difference can be found, second difference and it continues till stationary conditions are satisfied.

The first difference of a time series can be written as follows:

$$\nabla x_t = x_t - x_{t-1}. \tag{3.6}$$

The second difference of a time series can be written as follows:

$$\nabla x_t = x_t - x_{t-1} - x_{t-2}. \tag{3.7}$$

The pattern may continue until the time series is stationary.

3.9 Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

ACF and PACF are measure of relationship among present and previous series values and shows which previous series value are most convenient in predicting future values.

3.9.1 Autocorrelation Function (ACF)

The autocorrelation function (ACF) for a time series X_t, t=1,2,...,T, is the sequence ρ_k, k=0,1,2,...,N-1. ACF refers to the correlation of time series with past and future values. Autocorrelaation function is also called "lagged correlation" or "serial correlation". which refers to the correlation between members of a time series of numbers arranged in time. Autocorrelation function is used to determine the moving average for an autoregressive integrated moving average (ARIMA) model, (Ihaka, 2005; Montgomery et al., 2015).

Let X_t be stationary time series with mean μ, variance σ^2 and autocovariance γ_k then the autocorrelation function is given by:

$$\rho_k = \frac{cov(x_t, x_{t-k})}{\sqrt{var(x_t)var(x_{t-k})}} = \frac{\gamma_k}{\gamma_0}, \tag{3.8}$$

where γ_0 is the variance of the series and ρ_0=1.

i.The autocorrelation function is an even function of the lag such that: ρ_k=ρ_{-k}.

ii.$|\rho_k|\leq 1$.

3.9.2 Partial Autocorelation Function (PACF)

Partial autocorrelation function is a conditional correlation, it is a correlation between two variable X_t and X_{t-k} after removing any linear independence on $X_1, X_2, ..., X_{t-k-1}$ (Adhikari and Agrawal, 2013; Ihaka, 2005; Montgomery et al., 2015). The partial lag-k autocorrelation is denoted by Φ_{kk}, k=1,2,..., n-1.

Let X_t be a stationary time series with E(X_t)=0. The partial autocorrelation coefficient is defined as:

$$\Phi_{kk} = Corr(X_t, X_{t-k}|X_{t-1}, X_{t-2}, ..., X_{t-k+1}). \tag{3.9}$$

3.10 Time series models

To be able to analyse or make smooth inference about the data generating progress it is important to draw simple and reasonable statements about the procedure. A characteristic feature of time series data which differentiate the type of data is that the observations are generally correlated or depended. These are some technique that are to be used to model the stationary time series model such as: Autoregressive (AR) process and Moving average (MA)process models.

3.10.1 Autoregressive process

Autoregressive (AR) process is a statistical forecasting model in which predicted values are calculated only based on previous values of a time series letting z_t be a purely random process, then x_t is said to be AR process (model) of order p, which denotes the AR(p) model and it is given by:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + Z_t, \tag{3.10}$$

where X_t is the time series and Z_t is white noise with mean zero and variance σ_t^2, (Ihaka, 2005).

3.10.2 Moving Average process

The time series X_t is a moving average process of order q or MA (q) process if,

$$X_t = Z_t + \theta_1 Z_{t-1} + ... + \theta_p Z_{t-q}, \tag{3.11}$$

where $Z_t \sim WN(0, \sigma^2)$ and $\theta_1, \theta_2, ..., \theta_q$.

Hence WN-means white noise is a very important example of a stationary process.

The MA(q) process can also be written in the following equal form,

$X_t = \theta(B)Z_t,$

where moving average operator $\theta(B)$=1+$\theta_1 B + \theta_2 B^2 + ... + \theta_q B^q$

defines a linear combination of values in the shift operator $B^k Z_t = Z_{t-k}$.

3.10.3 Autoregresssive Moving Average process

Autoregressive Moving Average (p,q) process is a time series model of order p and q denoted by ARMA(p,q) is the combination of AR (p) and MA (q) model. In an AR (p) process the future value of a variable is assumed to be a linear combination of p past observations with a random error and a constant term. The mathematical formula of ARMA (p,q) model is defined as:

$$X_t = \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + Z_t - \theta_1 Z_{t-1} - ... - \theta_q Z_{t-q}, \tag{3.12}$$

where X_t represent the time dependent series, ϕ, i=1,2,...,p are nonseasonal AR parameters, θ, i=1,2,...,q are the nonseasonal MA parameters. An autoregressive model estimates values for the dependent variables, Z_t, as a regression function of previous values $X_{t-1}, X_{t-2}, ..., X_{t-p}$, autoregressive model has been applied extensively in hydrology for annual and periodic hydrologic time series. A moving average model is conceptually a linear regression of the current value of the series against the white noise or random shocks of one or more prior values of the series(Taheri et al., 2014).

3.10.4 Seasonal Autoregressive Moving Average progress

The Seasonal Autoregressive Moving Average (SARMA) of order (p,q) and (P,Q) is denoted by SARMA $(p, q) \times (P, Q)$ is defined by :

$$\Phi_P(B^S)\phi_p(B)x_t = \Theta_Q(B^S)\theta_q(B)z_t, \tag{3.13}$$

where S is the seasonality,

$\Phi_P(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - ... - \Phi_P B^{PS}$,

$\phi_p(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - ... - \phi_p B^p$,

$\Theta_Q(B^S) = 1 - \Theta_1 B^1 - \Theta_2 B^{2S} - ... - \Theta_Q B^{QS}$ and

$\theta_q(B) = 1 - \theta_1 B^1 - \theta_2 B^2 - ... - \theta_q B^q$.

3.10.5 Autoregressive Integrated Moving Average process

The autoregressive integrated moving average (ARIMA) process is a forecasting technique that predicts the future values of time series histrorical data and a series of errors' linear combination. The ARIMA model is given by:

$$X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}. \tag{3.14}$$

The left-hand side of the equation represent Autoregressive AR (p), and the right-hand side represent the Moving Average MA (q), (Delima, 2019).

3.10.6 Seasonal Autoregressive Integrated Moving Average process

The Seasonal Autoregressive Integrated Moving Average (SARIMA) process is an extension of ARIMA that explicitly supports univariate time series data with seasonal component. The SARIMA model attempts to capture the seasonal and nonseasonal relationship among the successive observations in a time series through sequences of ordinary as well as seasonal differencing of the series. The model is denoted by SARIMA $(p, d, q) \times (P, D, Q)^s$ model. The parameter (p,P), (q,Q) represent the autoregressive and moving average proessess while (d,D) specify the degrees of ordinary and seasonal differencing (Adhikari and Agrawal, 2013; Arbain and Wibowo, 2012).

SARIMA model is defined by:

$$\Phi_P(B^s)\phi_p(B)X_t = \Theta_Q(B^s)\theta_q(B)Z_t, \tag{3.15}$$

where B is the lag defined by$Bx_t = x_{t-1}$, $\Phi_P, \phi_p, \theta_q, \Theta_Q$ are the lagged polynomials in B of order p, P, q and Q.

Z_t is a series of purely random error and X_t is the stationary nonseasonal series which is obtained after the differencing process.

3.10.7 Vector Autoregressive process

The vector autoregressive (VAR) model is a multivariate regression technique in which each dependent variable is regressed on lags of itself and on lags of all other dependent variables in the model. The VAR model will be used to detect the statistical relationship between water evaporation and temperature also water evaporation and rainfall. The VAR model is defined by:

$$M_t = K + A_1 M_{t-1} + A_2 M_{t-2} + ... + A_n M_{t-n} + \varepsilon_t, t = 0, \pm 1, \pm 2..., \quad (3.16)$$

where $M_t = [m_i t, ..., m_k t]'$ is $(n \times 1)$ random vector, the A_i are fixed $(n \times n)$ coefficient matrices, K is a $n \times 1$ vector of constants allowing for the non-zero mean E(μ_t), (Issaka, 2015; Shahidi et al., 2020).

3.10.8 Autoregressive Conditional Heteroscedasticity

The autoregressive conditional heteroscedasticity (ARCH) model is a statistical model for time series data that describes the variance of current error terms, this model was first presented in economic studies by Engle (1982). The ARCH model is a good fit when the error variance in a time series follows an autoregressive (AR) model; if an autorgressive moving average (ARMA) model is assumed for error variance, the model is a generalized autoregressive conditional heteroscedasticity (GARCH) model (Zhao et al., 2021).

The ARCH model is defined by:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{b} a_i \varepsilon_{t-i}^2, \varepsilon_t = \sigma_t z_t, \quad (3.17)$$

where σ_t^2 is the conditional variance, ε_t is the error term of the model with mean μ_t=0 and variance σ_t^2=1.

3.10.9 Generalized Autoregressive Conditional Heteroscedasticity

The generalized autoregressive conditional heteroscedasticity (GARCH) model is a privileged approach for modelling volatility. The GARCH model will be used to determine the degree of variation of water evaporation over time.

The GARCH model for a process (ε_t) is defined as :

$$\sigma_t^2 = var(\varepsilon_t/\varepsilon_u, u < t) = w + \sum_{i=1}^{u} \alpha_i \varepsilon_{t-1}^2 + \sum_{j=1}^{m} \beta_j \sigma_{t-j}^2, \tag{3.18}$$

$\varepsilon = \sigma_t e_t, e_t \sim Normal(0, 1)$,

$\varepsilon_t/\varphi_{t-1} \sim Normal(0, \sigma_t^2)$,

where $\sigma_t^2(var/\varepsilon_t/\varepsilon_u, u\varepsilon t)$ is the conditional time-variate variance of the residual series, φ is a constant, $\alpha_1, \alpha_2, ..., \alpha_u$ and $\beta_1, \beta_2, ..., \beta_m$ are the coefficient of the GARCH (u,m) and u and m are the orders of the coefficients, (Francq et al., 2011; Montgomery et al., 2015).

3.11 Testing for non-stationary time serie process

There are three tests used to test for non-stationarity namely: unit root test, Dickey-Fuller test and Augmented Dickey Fuller.

3.11.1 Unit root test

The unit root test is mainly a descriptive tool performed to classify if a series is stationary or non-stationary. If a data series appears to be non-stationary, think as the maintained hypothesis that it is non-stationary and integrated. The hypothesis will be rejected only and only if there is a clear evidence for rejection. An AR model will be used where AR(1): X_t=$\phi X_t -_1 + Z_t$ to apply

the unit root test.

The hypotheses are:

H_0: ϕ=1 (Test has unit root)

$H_1 : |\phi| < 1$ (Test is stationary)

The test statistic is given by:

$$t - stat = \frac{\widehat{\phi} - 1}{SE(\widehat{\phi})}, \tag{3.19}$$

where $(\widehat{\phi})$ is the estimate of the least square and $SE(\widehat{\phi})$ is the standard error of the estimate. We reject the hypothesis if $t - stat > t_{critical}$.

3.11.2 Dickey-Fuller test

Dickey Fuller test is a test that determines whether a process has a unit root test or not. The AR (1) model will be used.

$X_t = \phi x_{t-1} + z_t$

$\nabla X_t = x_t - x_{t-1}$

$=> \nabla X_t = \phi_1 x_{t-1} + z_t - x_{t-1}$

$=> \nabla X_t = (\phi_1 - 1)x_{t-1} + z_t$

Let $\beta = \phi - 1$, then $\nabla X_t = \beta x_{t-1} + z_t$.

The hypotheses are:

$H_0 : \beta = 0$ (This is equivalent to $\phi = 1$)

$H_1 : \beta < 0$ (This is equivalent to $\phi < 1$)

Test statistic is given by:

$$\tau_{stat} = \frac{\widehat{\beta}}{SE(\widehat{\beta})}, \tag{3.20}$$

where $SE(\widehat{\beta})$ will be the standard error of the coefficient estimate. The null hypothesis is rejected if the test statistic is more negative than the critical value(Mushtaq, 2011).

3.11.3 Augmented Dickey-Fuller test

The Augmented Dickey-Fuller(ADF) test is developed by David Dickey and Wayne Fullr in 1979. The ADF test allows testing the presence of a unit root test in intended time series based on considering an autoregressive AR (1) model. If the autoregressive parameter of the AR (1) model is equal to one the intended time series is considered a non-stationary process (null hypothesis), otherwise it is considered stationary (Fathian, 2019; Mushtaq, 2011).

The hypotheses are:

$H_0 = \beta = 0$

$H_1 = \beta < 0$

The test statistic is given by:

$$\tau_{stat} = \frac{\widehat{\beta}}{SE(\widehat{\beta})}. \tag{3.21}$$

The test statistic of ADF is compared with the critical value of the ADF test, if the test statistic is less than the critical value, the test rejects the null hypotheses and conclude that there is no unit root test.

3.12 Box-Jenkins technique

Box-Jenkins technique is a systematic method of identifying, fitting, checking and using integrated autoregressive moving average (ARIMA) time series

models, it modifies the time series to make it stationary by using deference between data points. The Box-Jenkins technique is applied to find the appropriate model for time series to be used, the model can also make forecasts of future economic activity based on the past activities, (Babazadeh and Shamsn, 2014).

Box-Jenkins technique is also known as ARIMA. This is the most important method to evaluate the stationary of time series. There are three steps in the Box-Jenkins technique process, the main steps can be used as many times as possible.

The Box-Jenkins technique steps are as follows:

i. Model specification

ii. Model fitting

iii. Model diagnostic

3.12.1 Model specification

Model specification is the process of developing a regression model. The process of model specification consists of selecting an appropriate functional form for the model and choosing which variable to include. A good model identification has a selected model that is statistically sound and practically meaningful. A parsimonious model is frequently preferred when two applicant models seem to be equally good. Parsimonious model is a model that accomplishes an anticipated level of clarification or perdition with several predictor variables as possible, (Lancsar et al., 2017; Perera and Silvapulle, 2018).

3.12.2 Model fitting

Model fitting consists of finding the best possible estimates of those unknown parameters within a given model. After identifying the model and estimating the unknown parameters, the model needs to be checked for goodness of fit, this is done through model diagnostics, .

3.12.3 Model diagnostic

The main aim of model diagnostics is to analyse the quality of the model that has been identified and estimated, model diagnostic is concerned with testing the goodness of fit of a tentative model. In this section some standard methods for model diagnostics are outlined. The most frequently used techniques are residual derived from fitted model behave like white noise process, (Bernal et al., 2018) .

3.12.4 Residual analysis

Residual analysis is the difference between observed value of the dependent variable (X) and the predicted value ($\widehat{X}$), before the model should be assessed using residuals, (Martin et al., 2017). Residuals are given by:
Residual=Actual value-predicted value.

A good model is the one with residuals that satisfy the following properties:

i. Independent (uncorrelated error)

ii. Normality distribution

iii. Constant variance.

Test for normality

Normality test can be used to determine if a data set is well-modelled by a normal distribution and to estimate how possible it is for a random variable causing the data set to be normally distributed, (Ghasemi and Zahediasl, 2012) .

Test of normality can be performed by:

Constructing a histogram: Gross normality can be assessed by plotting histogram of residuals. Histogram of normally distributed residuals should be symmetric and bell shape.

Test of constant variance

Test of constant variance can be inspected by plotting the residuals over time. If the model is adequate the plot suggest a regular scatter around a

zero-horizontal level with no trends.

Test of independence

A test of independence can be performed by:

1. Examining the ACF of the residuals

To compute the sample ACF (autocorrelation function) of the residuals, a time series data must be used to determin a pattern of ACF residual plot. If the residual do not form any pattern and are also statistically insignificant there are independent, that means they are within $Z_{\frac{\alpha}{2}}$ standard deviation.

2. Using run test

A run-test is a statistical technique that look at a string of data that is occurring randomly from a definite distribution. The run test examines the occurrence of related events that are divided by events that are different. Different procedure to test the independence in residuals can be used. The runs test examines the residuals sequence to look for patterns. If the residuals are correlated, the pattern will occur.

Residual ACF and PACF plots

A residual plot is a graph that represent the residuals on the vertical axis and the independent variable is on the horizontal axis. ACF plots are commonly used tool checking randomness in s time series data. These plots are used in the model identification stage for Box-Jenkins autoregressive and moving average time series process. Data that have the significant autocorrelation is not random.PACF plot a are useful in identifying the order of an Autoregressive process. The PACF of an AR (p) process is zero at lag k+1 and greater, (Mohanasundaram et al., 2013).

3.12.5 Ljung-Box Statistics

The Ljung-Box (1978) test is used to test the lack of fit of a time series model. The test is applied to the residuals of a time series after fitting an ARMA (p,q) process. The Ljung-Box test is based on the autocorrelation plot, instead of testing randomness at each distinct lag, it test the overall randomness based on number of lags,(Brownlee, 2017; Burns, 2002).

Hypotheses are:

H_0 : the process is white noise

H_1 : the process is not white noise

Test statistic is given by:

$$Q = n(n+2)\sum_{i=1}^{k}\frac{\widehat{\rho}}{n-1}, \tag{3.22}$$

where ρ_1 is the ith PACF. The null hypothesis is rejected if $Q > X^2_{k-p-q}(\alpha)$.

3.12.6 Quantile-Quantile Plots

The quantile-quantile (Q-Q) plot is a graphical method that is used for determining if two data sets comes from the same distribution. It is a scatter plot created by plotting a set of quantiles of the first dataset against quantiles of the second data set. If the two graphs comes from the same distribution the Q-Q plots should fall along the straight line, (Lai, 2017).

3.13 Model selection

Model selection is the process of selecting a statistical model from a set of candidate models. Several appropriate models may be used to select a model to analyse time series to present a given set of data. Numerous criteria are introduces to compare model which are different from methods for model recognition. The study will look at Principle of Parsimony, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The model with the smallest AIC and BIC value will be selected as the appropriate model, (Montgomery et al., 2015).

3.13.1 Principle of Parsimony

The principle of parsimony is a model that achieves a level of goodness of fit using as few variables as possible, this means the simplest model should be chosen. Parsimony is one of the principle of reasoning in science. Parsimony principle was named by William of Ockham, a 14th century logician and Franciscan monk who used this principle in his philosophial reasoning.

Principle of parsimony is also defined as the problem of statistical modelling that has an objective function which minimize the complexity subject to constraint of the model suitability (Vandekerckhove et al., 2014).

3.13.2 Akaike Information Criterion

Akaike Information Criterion (AIC) was first developed by Akaike (1973) as a way to compare different models on a given outcome. AIC is a fined method based on in-sample fit to evaluate the likelihood of a model to predict or estimate the future value. The AIC rewards goodeness of fit and includes penalty that is an increasing function of the number of estimated parameters. The penalty does not encourage ovrerfitting of the model, because increasing the number of parameters in the model almost improves the goodness of fit,the AIC estimates water evaporation, then the model with the lowest AIC is selected to be the best model, (Dabral et al., 2017). The AIC is defined by:

$$AIC = -2lnL + 2 * k, \tag{3.23}$$

where k=(p+q+P+Q) is the number of terms estimated in the model,

L= denotes the likelihood, the measure of model fit.

3.13.3 Bayesian Information Criterion

Gideon E. Schwarz in (1978) developed a Baysian extension of minimum AIC. Bayesian Information Criterion (BIC) is another model selection that measures the trade off between model fit complexity of the model (Montgomery et al., 2015). BIC is defined by:

$$BIC = -2lnL + k * ln(n), \tag{3.24}$$

where L is the likelihood, n is te number of recorded measurement and k are the number of estimated parameters.

k= number of model parameter,

n = the number of data points in x, the number of observation (evaporation),

L= likelihoods a measure of model fit.

3.14 Parameter estimation

The parameter estimation is the process of using sample data to estimate the parameter of the selected distribution. There several parameter estimation methods such as the methods o moments, maximum likelihood, least squares that is to be used to estimate the parameters in the identified model, (Montgomery et al., 2015).

3.14.1 Methods of Moments

The method of moments is a technique for estimating the parameters of a statistical model. The methods of moments it finds the values of the parameters that result in a match between the sample moments and the population moments. Moments method consists of equating sample moments such as the sample mean $\bar{X}$, sample variance and sample autocorrelation function to the theoretical counters and solving the resultant equation, (Hazelton, 2011),

Estimating $E(X_t)$

The natural estimator for the mean $\mu = E(X_t)$ of the stationarity process is the sample mean given by:

$$\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t. \tag{3.25}$$

3.14.2 Least squares estimator

The least square estimator is a statistical method of estimating values from a set of observations by minimizing the sum of the square of the different among the observation and the values to be found, (Everitt and Howell, 2021).

The method of least square estimator (LSE) is an estimation procedure developed for standard regression models. For simple linear regression the model is given by:

$$X_t = \alpha + \beta W_t, t = 1, 2, ..., n. \tag{3.26}$$

The least squares estimate is given by:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(X_t - \widehat{X})(W_t - \widehat{W})}{\sum_{i=1}^{n}(W_t - \widehat{W})^2}, \tag{3.27}$$

and $\widehat{\alpha} = \widehat{Y} - \widehat{\beta}\widehat{W}$.

3.14.3 Maximum likelihood methood

The maximum likelihood method (MLE) is a technique of estimating the parameters of a model. This estimation method is widely used method. The maximum likelihood selects the set of values of the model parameters that maximize the likelihood function, it gives a unified approach to estimation, (Hurlin, 2013).

Let $X_t = (X_1, X_2, ..., X_n)$ be a vector of stationary time series and $X_1, X_2, ..., X_n$ be a vector of original observation.

The ARIMA (p,d,q) model can be expressed as:

$$W_t = \phi_1 W_{t-1} + ... + \phi_q W_{t-q} + Z_t + \theta_1 Z_{t-1} + ... + \theta_p Z_{t-p}. \tag{3.28}$$

The MLE is also the procedure of estimating the value of one or more parameters for a given statistic. The MLE of parameter θ can be written as $\widehat{\theta}$:

$L(\theta|X) = f(x|\theta)$,

where $\theta \epsilon \Theta$ and $X = (x_1, x_2, ..., x_n)$.

3.15 Time series forecasting

Time series forecasting is a technique for prediction of events through a sequence of time, the technique predicts future events by analysing the trend of the past. Time series forecasting is an integral part of decision making, as it can play a key role in many areas. Modern organization requires short-term, medium-term and long-term forecasting depending on the specific application, (Montgomery, 2008).

3.15.1 Short-term forecasting

Short-term forecasting is the prediction of events for only a few time periods (days, weeks, and months) into the future. For example, forecasting price change in business and weather, (Montgomery et al., 2015).

3.15.2 Medium-term forecasting

Medium-term forecasting extend beyond from one to two years into the future. In most cases the medium forecasting is more convenient to determine and identify the patterns that are found in the historical data, (Montgomery et al., 2015).

3.15.3 Long-term forecasting

Long-term forecasting is the prediction of the future values that extend to more than two perhaps 5 five years and above. Since the historical data does not change quickly, the more profound forecasting method is short-term and medium-term. Short-term and medium-term forecasting are used to predict the future values and to model the patterns that are found in the historical data, (Montgomery et al., 2015).

3.16 Time series forecasting methods

There are different methods that can be used to model time series forecasting. These forecasting methods are appropriate for forecasting data with no trends or seasonal pattern. The time series forecasting methods to be used are: ARIMA, MA, weighted moving average (WMA) and exponential smoothing.

3.16.1 ARIMA forecasting

An (ARIMA) model is a class of statistical models for analysing and forescating time series data. ARIMA is usally superior to exponential smoothing techniques when the data is reasonably long and the correlation between past observations is stable. The AR (p) model takes the p lags of the forecasting errors to improve the forecast, an AR (p) model is defind as:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \varepsilon_t, \tag{3.29}$$

where X_t and ε_t are the actual values and the random error at time t,(Bakar and Rosbi, 2017).

3.16.2 Moving average

Moving average (MA) is a technique to get an overall idea of the trend in a data set. The MA method is the forecasting method that discards all observation that are more than k units by using the average of the most recent k data values in the time series as the forecast for the next period, and the weight to each of the k most recent observations is equal and constant over time.

To begin forecasting the future values, k observations are needed to compute the new forecasts, all k of the most recent observations must be retained from step to step.

The MA of the length k is given by:

$$\hat{X}_{t+1} = \frac{\sum_{t=1}^{n} X_{t-k+1}}{k}, \tag{3.30}$$

where $\hat{X}_{t+1}$ is the forecasting of the time series for period t+1.

3.16.3 Weighted Moving Average

The weighted moving average (WMA) is a technical indicator that assigns a greater weighting to the most recent data points and less weighting on past data points. This is used for smoothing irregular fluctuations in a time series to permit the data analyst to better reveal trend/ cycle pattern over time. Most of the time WMA is used to compute short-term forecasts of time series, (Perry, 2010). The average of the observations used to compute the forecast is found by multiplying each data by weighting factor k. The WMA length k is given by:

$$\hat{X}_{t-1} = \sum_{i=0}^{k-1} W_i X_{t-k}, \tag{3.31}$$

where W_i=weight to the $(t-1)^{th}$ observation i= 0, 1, 2,..., (k-1).

3.16.4 Exponential Smoothing

Exponential smoothing is the statistical procedure for identifying significant changes in data by ignoring the variations unrelated to the purpose at hand. Exponential smoothing is one of the forecasting method that used weighted average of a past time series values as a forecast. This method is known to be the method used for several reasons; including appeal, ease in updating small amount of information that must be carried from one person to the next. In the exponential smoothing old data is given less relative weight than the new data which is given more relative weight, (Ostertagova and Ostertag, 2011). The exponential method is given by:

$$\widehat{X}_{t+1} = \beta X_t + (1 - \beta)\widehat{X}_t, \tag{3.32}$$

where

$\widehat{X}_{t+1}$ is the forecast value for time period X_t is the actual value of the time series in period t,

$\widehat{X}_t$ is the forecast of the time series for period t and

β is the smoothing constant $0 \leq \beta \geq 1$.

3.17 Measuring forecasting accuracy

The accuracy of forecasting is the degree of closeness of the statement of quantity to that quantity's true value. Mostly the true value cannot be determined at the time, the forecast is made because the statement relates to the future. Measuring forecasting accuracy is one greatest concerns of forecasting the future values of the time series. It is important to evaluate the aspect of any recommended forecast techniques. The accuracy when is made it provides a quantitative estimate of the expected quality of the forecasts, (Chaussonnet et al., 2013).

The standard measure of the forecast accuracy is the mean forecast error (MFE), the mean absolute deviation (MAD), the mean square error (MSE) and the mean absolute percent error (MAPE), will be used to the accuracy of the model.

3.17.1 Mean forecast error

Mean forecast error (MFE) is a measure of how accurate the forecast is in a given period. The forecast error is calculated by taking the diffrence between the real value and the predicted values of the time series and is given as follows:

$$e_t(l) = X_t - \widehat{X}_t, t = 1, 2, 3, ..., n. \tag{3.33}$$

Mean forecast error in one-time period does not show much information, the accumulation of error over time should be taken into consideration. The accumulation of the forecast error e_t value is not always revealing, for some of them will be positive errors and some will be negative. The average forecast error or mean forecast is given by:

$$MFE = \frac{\sum_{t=1}^{n}(X_t - \widehat{X}_t)}{n}. \tag{3.34}$$

3.17.2 Mean absolute deviation

The absolute deviation is the deviation in which only the size of the error is considered regardless of the sign of the error. Mean Absolute Deviation (MDA) measures is used to measure the accuracy of the prediction by averaging the alleged error (the absolute value of each error). If these absolute deviation are accumulated over time and find the average of these absolute deviation, (Khair et al., 2017; Swanson et al. , 2011), the measure of accuracy is referred to as the MAD and is given by:

$$MAD = \frac{\sum_{t=1}^{n}(X_t - \widehat{X}_t)}{n}. \tag{3.35}$$

3.17.3 Mean square error

Mean square error (MSE) is another way to eliminate the problem of positive errors cancelling the negative errors is to square the forecst error. Regardless of wheather the forecast error has a positive or negative sign, the square error will always be positive sign, (Khair et al., 2017; Swanson et al. , 2011). When the squared errors are accumulated over time and find the average of these errors,this is called MSE and is given by:

$$MSE = \frac{\sum_{t=1}^{n}(X_t - \widehat{X}_t)^2}{n}. \tag{3.36}$$

3.17.4 Mean absolute percent error

The mean absolute percent error (MAPE) is computed as average of the absolute difference between the forecast and actual values, expressed as a percentage of the actual value. The MAPE indicates how much error in predicting compared with the real value. The problem with both MAD and the MSE is that their values depend on the magnitude of the item being forecast. If the forecast item is measured in thousands or millions, the MAD and MSE values can be very large. To avoid this problem, MAPE can be used, (Khair et al., 2017; Swanson et al. , 2011).

The MAPE is given by:

$$MAPE = \frac{\sum_{t=1}^{n}(|X_t - \widehat{X}_t|)}{n} \times 100. \tag{3.37}$$

Chapter 4

Results and discussion

4.1 Introduction

This chapter describes the analysis of water evaporation rate using techniques discussed in chapter 3 and the techniques are: time series plots (trend, stationarity and auto-correlation functions), fitting the (ARIMA, ARH, GARCH and VAR) model residuals analysis and forecasting of the water evaporation rate from the selected dams in Limpopo province South Africa. A statistical packages used for analysis is SPSS, R softwear and Eviews.

4.2 Data Analysis

In this section the data presented in chapter 3 under the section of data source will be analysed. The time series plots pattern will be identified and discussed. If the pattern is non-stationary, differencing method is applied to make the time series pattern to be stationary.

4.2.1 Descriptive Statistics

Table 4.1: Summary descriptive statistics of water evaporation rate data.

Descriptive Statistics						
	Names of Dams	N	Minimum (mm)	Maximum (mm)	Mean (mm)	Std. Deviation
Evaporation	Mokolo Dam	399	0.48	67.43	4.16	1.91
	Ga-Rantho Dam	4092	0.18	29.97	3.76	1.99
	Leerkraal DeHoop Dam	4085	0.33	16.19	3.76	1.88
	Luphephe Dam	4089	0.14	33.91	3.17	1.92
Valid N(list wise)		16265				

Table 4.1 presents the summary descriptive for evaporation rate data. The results reveal the minimum, maximum, average and the standard deviation values for evaporation rate for the selected dams and they read as: Mokolo dam (min = 0.48, max = 67.43, mean = 4..16 and Sd = 1.91), Ga-Rantho dam (min = 0.18, max = 29.97, mean = 3.76 and Sd = 1.99), Leeukraal De-Hoop (min = 0.33, max = 16.19, mean = 3.76 and Sd = 1.88) and Luphephe dam (min = 0.14, max = 33.91, mean = 3.17 and Sd = 1.92). The standard deviation shows how spread the data is from the mean. The standard deviation of the selected dams are small this indicates that the sample mean is more accurate of the evaporation mean.

4.2.2 Time series Analysis

Figure 4.1 presents a partial of non-stationary time series plot of water evaporation rate for Mokolo dam since the mean changes over time and the variance is not constant. It shows a slight increasing trend of water evaporation rate. The highest water evaporation rate is around the 25^{th} of every month of each year recorded and the lowest water evaporation rate was around the 17^{th} of every month of each year recorded. The time series plot of water evaporation increases due to its factors temperature and rainfall. The time series needs to be differenced to be perfectly stationary, since the mean and variance are slightly not constant. This implies that there is seasonality change effect (Shumway et al., 2000).

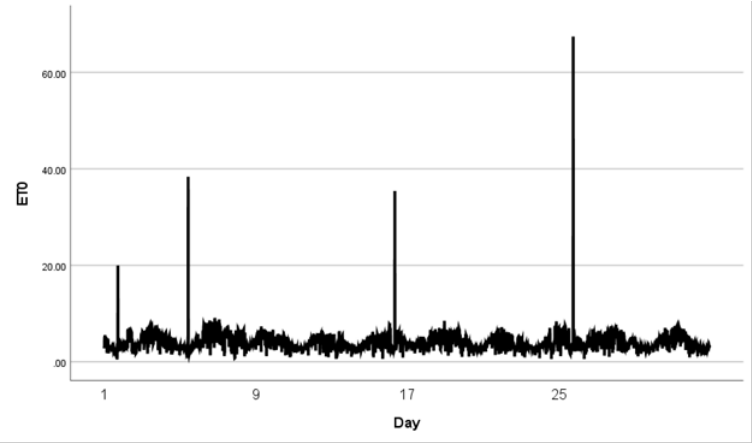


Figure 4.1: Time series plot of water evaporation rate (Mokolo Dam).

Figure 4.2 presents a non-stationary time series plot of water evaporation rate for Ga-Rantho dam since the mean changes over time and the variance is not constant. It shows an increase and a slight constant trend of water evaporation rate. The highest water evaporation rate is around the end of every month of each year recorded and the lowest water evaporation rate was around the 17^{th} of every month of each year recorded. The time series plot of water evaporation increases and decreases due to its factors temperature and rainfall. The time series needs to be differenced to be make it stationary, since the mean and variance are not constant. This suggests that there is seasonality change effect (Shumway et al., 2000).

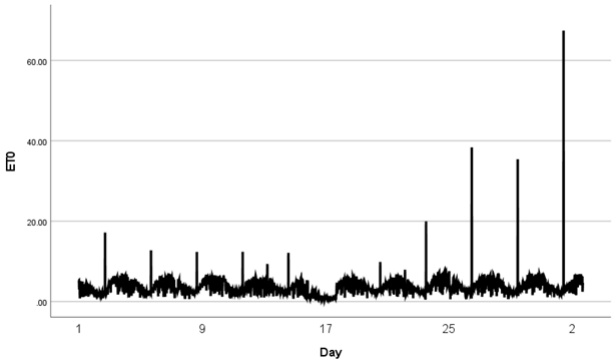


Figure 4.2: Time series plot of water evaporation rate (Ga-Rantho Dam).

Figure 4.3 presents a non-stationary time series plot of water evaporation rate for Leeukraal DeHoop dam since the mean changes over time and the variance is not constant. It shows an increasing and a decreasing trend of water evaporation rate. The highest water evaporation rate is around the 26^{th} of every month of each year recorded and the lowest water evaporation rate was around the 7^{th} of every month of each year recorded. The time series plot of water evaporation increases and decreases due to its factors temperature and rainfall. The time series needs to be differenced to make it stationary, since the mean and variance are not constant. This suggests that there is seasonality change effect (Shumway et al., 2000).

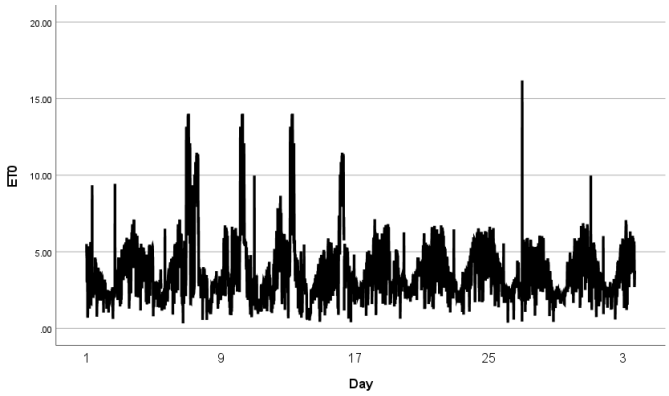


Figure 4.3: Time series plot of water evaporation rate (Leeukraal DeHoop Dam).

Figure 4.4 presents a non-stationary time series plot of water evaporation rate for Luphephe dam since the mean changes over time and the variance is not constant. It shows an increasing and decrease trend of water evaporation rate. The highest water evaporation rate is around the 26^{th} of every month of each year recorded and the low est water evaporation rate was around the 7^{th} of every month of each year recorded. The time series plot of water evaporation increases and decreases due to its factors temperature and rainfall. The time series needs to be differenced to be make it stationary, since the mean and variance are not constant. This present that there is seasonality change effect (Shumway et al., 2000).

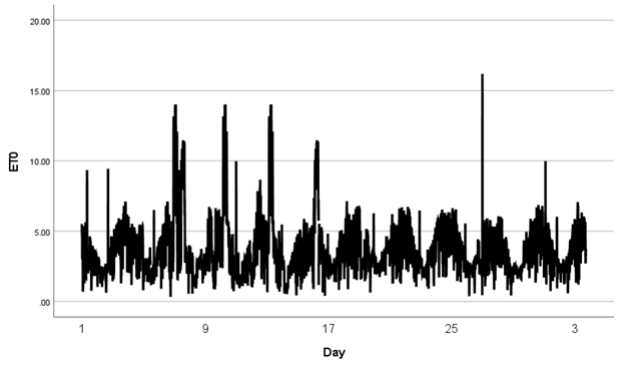


Figure 4.4: Time series plot of water evaporation rate (Luphephe Dam).

Figure 4.5 presents the autocorrelation function plot of water evaporation rate where the process is non-stationary. The peak of the plot starts from a high point and constantly decreases, all the peaks are significant. The ACF plot hence present an ARIMA model. The ACF plot suggested MA (1), MA (2), MA (3) and SMA (0) models (Nau, 2014).

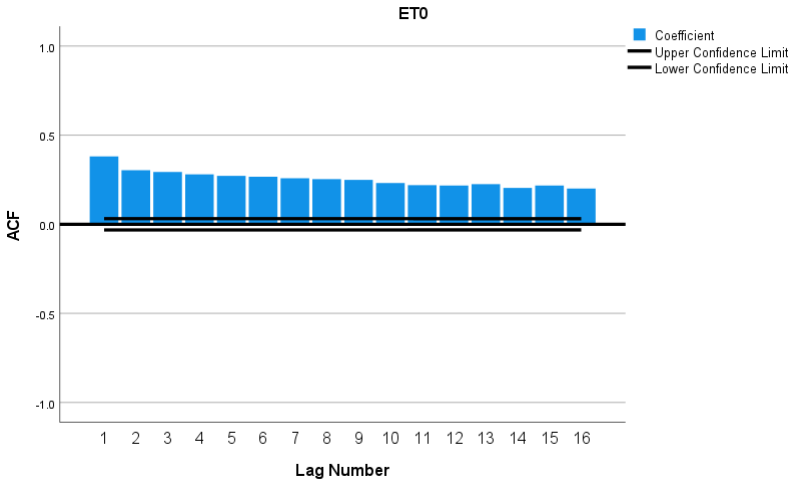


Figure 4.5: Autocorrelation Function of water evaporation rate (Mokolo Dam).

Figure 4.6 presents the autocorrelation function plot of water evaporation rate where the process is non-stationary. The peak of the plot starts from a high point and constantly decreases, all the peaks are significant. The ACF plot hence present an ARIMA model. The ACF plot suggested MA (1), MA (2), MA (3) and SMA (0) models (Nau, 2014).

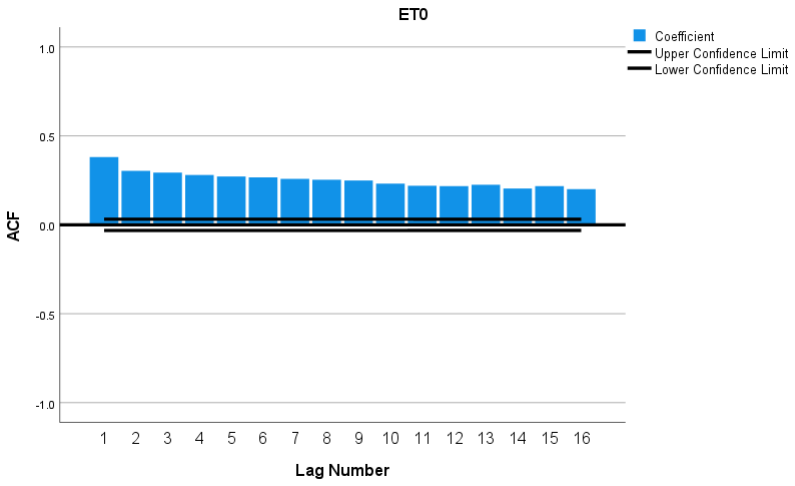


Figure 4.6: Autocorrelation Function of water evaporation rate (Ga-Rantho Dam).

Figure 4.7 presents the autocorrelation function plot of water evaporation rate where the process is non-stationary. The peak of the plot starts from a high point and constantly decreases, all the peaks are significant. The ACF plot hence present an ARIMA model. The ACF plot suggested MA (1), MA (2), MA (3) and SMA (0) models (Nau, 2014).

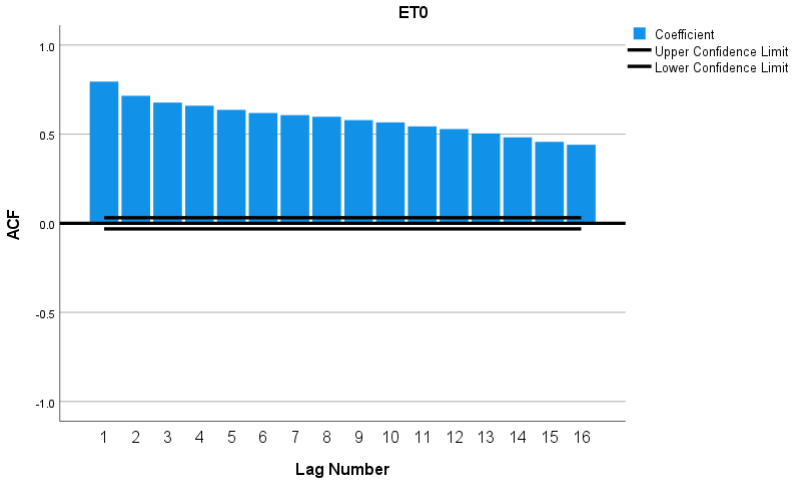


Figure 4.7: Autocorrelation Function of water evaporation rate (Leeukraal DeHoop Dam).

Figure 4.8 presents the autocorrelation function plot of water evaporation rate where the process is non-stationary. The peak of the plot starts from a high point and constantly decreases, all the peaks are significant. The ACF plot hence present an ARIMA model. The ACF plot suggested MA (1), MA (2), MA (3) and SMA (0) models (Nau, 2014).

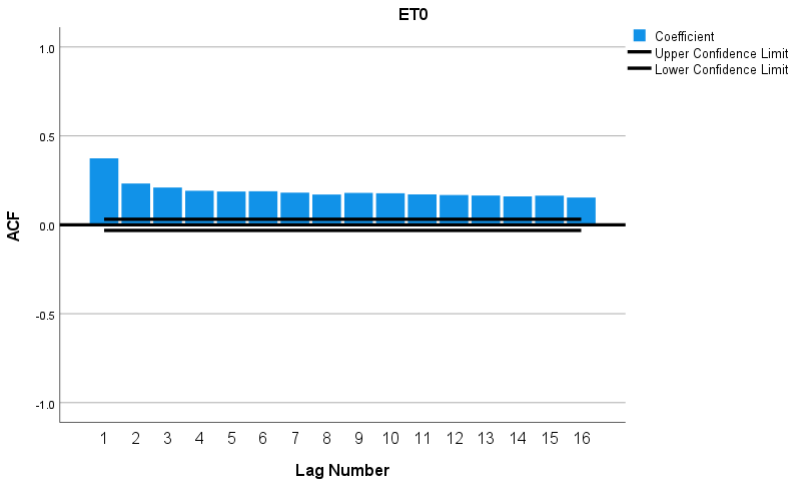


Figure 4.8: Autocorrelation Function of water evaporation rate (Luphephe Dam).

Figure 4.9 presents the partial autocorrelation function plot from lag 1 to lag 16 and it consist of non-stationary process. The PACF plot shows that from lag 1 to lag 6 the peaks are significant, hence the is no seasonality since the other peaks are insignificant. The PACF plot suggests AR(1), AR(2) and AR(3) mode (Nau, 2014).

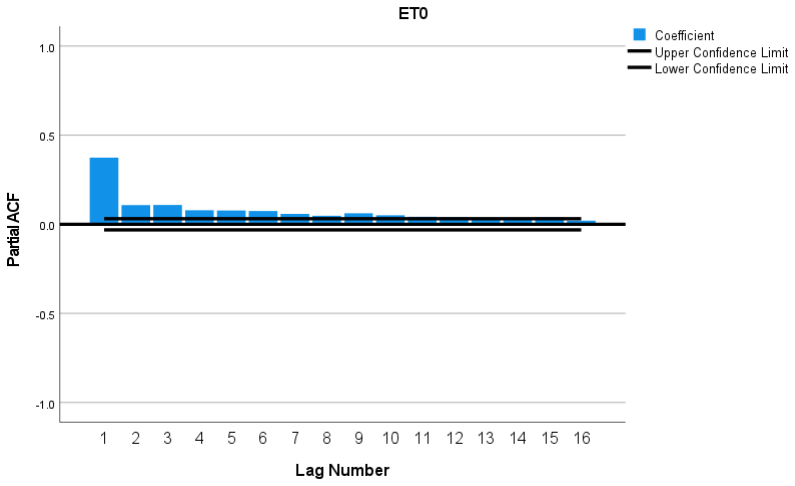


Figure 4.9: Partial Autocorrelation Function of water evaporation rate (Mokolo Dam).

Figure 4.10 presents the partial autocorrelation function plot from lag 1 to lag 16 and it consist of non-stationary process. The PACF plot shows that from lag 1 to lag 7 the peaks are significant, hence the is no seasonality since the other peaks are insignificant. The PACF plot suggests AR(1), AR(2) and AR(3) model (Nau, 2014).

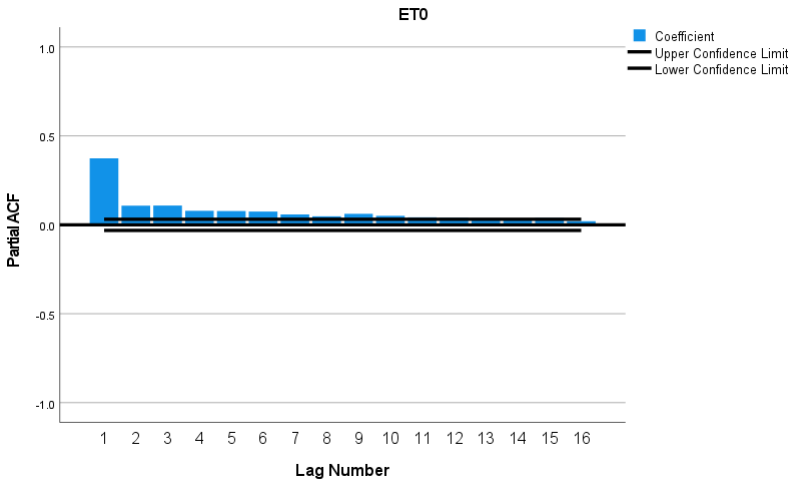


Figure 4.10: Partial Autocorrelation Function of water evaporation rate (Ga-Rantho Dam).

Figure 4.11 presents the partial autocorrelation function plot from lag 1 to lag 16 and it consist of non-stationary process. The PACF plot shows that from lag 1 to lag 10 the peaks are significant, hence the is no seasonality since the other peaks are insignificant. The PACF plot suggests AR(1), AR(2) and AR(3) model (Nau, 2014).

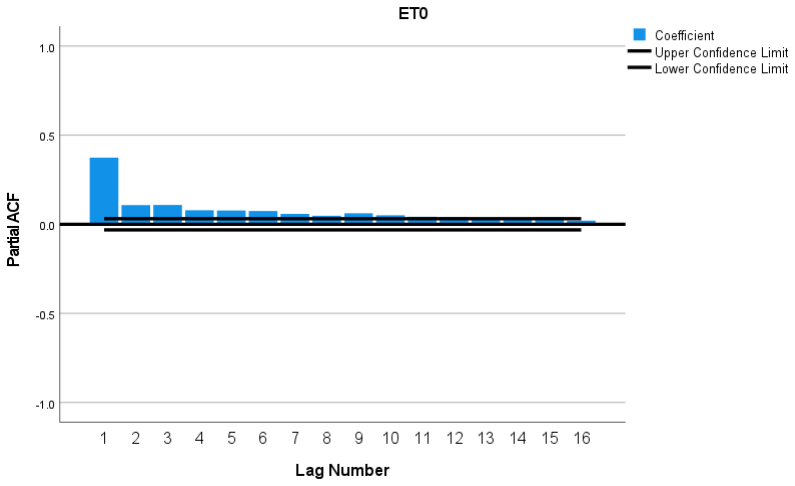


Figure 4.11: Partial Autocorrelation Function of water evaporation rate (Leeukraal DeHoop Dam).

Figure 4.12 presents the partial autocorrelation function plot from lag 1 to lag 16 and it consist of non-stationary process. The PACF plot shows that from lag 1 to lag 6 the peaks are significant, hence the is no seasonality since the other peaks are insignificant. The PACF plot suggests AR(1), AR(2) and AR(3) model (Nau, 2014).

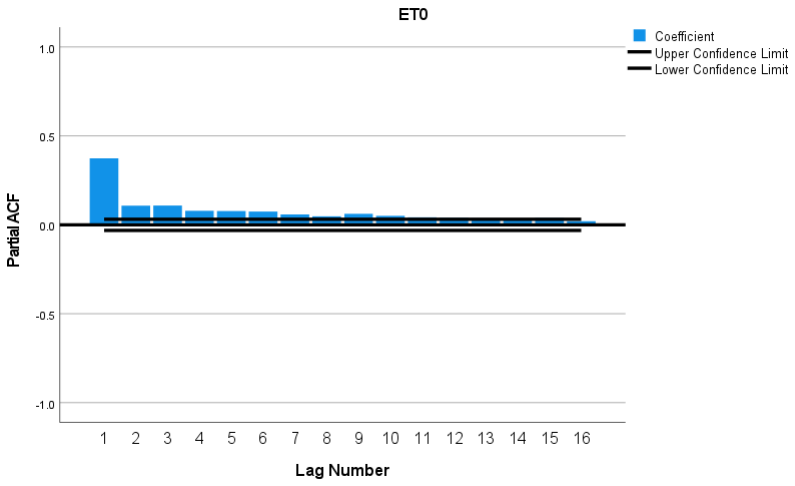


Figure 4.12: Partial Autocorrelation Function of water evaporation rate (Luphephe Dam).

Figure 13 presents that the first differenced time series plot resulted to be stationary after differencing. The water evaporation rate trend shows an accurate pattern, where the mean and the variance of the water evaporation rate time series is constant, hence there is no need for second differencing (Adhikari and Agrawal, 2013; Shumway et al., 2000).

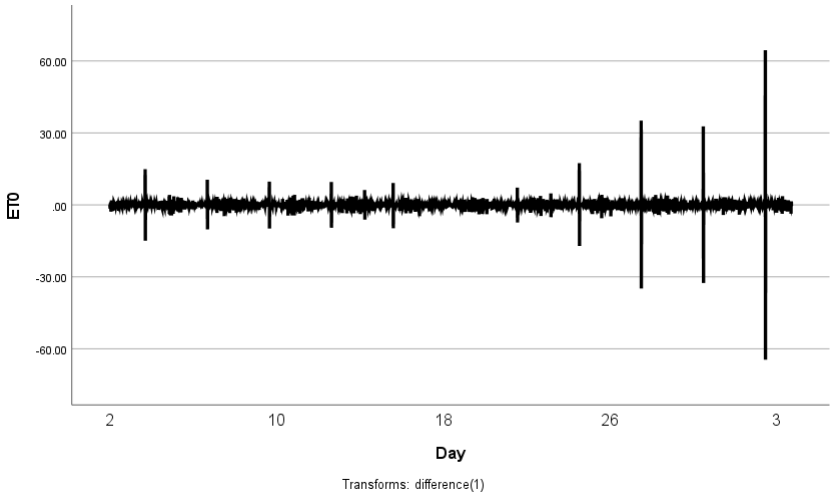


Figure 4.13: First difference time series plot of water evaporation rate (Mokolo Dam).

Figure 4.14 presents that the first differenced time series plot resulted to be stationary after differencing. The water evaporation rate trend shows an accurate pattern, where the mean and the variance of the water evaporation rate time series is constant, hence there is no need for second differencing (Adhikari and Agrawal, 2013; Shumway et al., 2000).

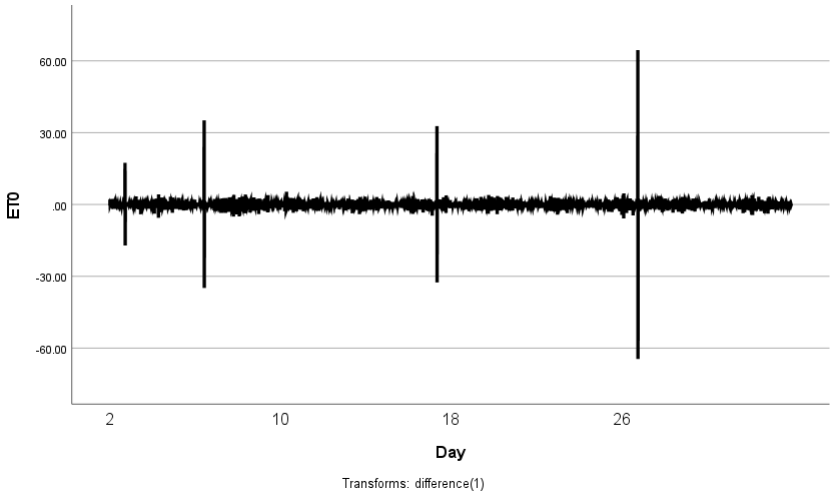


Figure 4.14: First difference time series plot of water evaporation rate (Ga-Rantho Dam).

Figure 4.15 presents that the first differenced time series plot resulted to be stationary after differencing. The water evaporation rate trend shows an accurate pattern, where the mean and the variance of the water evaporation rate time series is constant, hence there is no need for second differencing (Adhikari and Agrawal, 2013; Shumway et al., 2000).

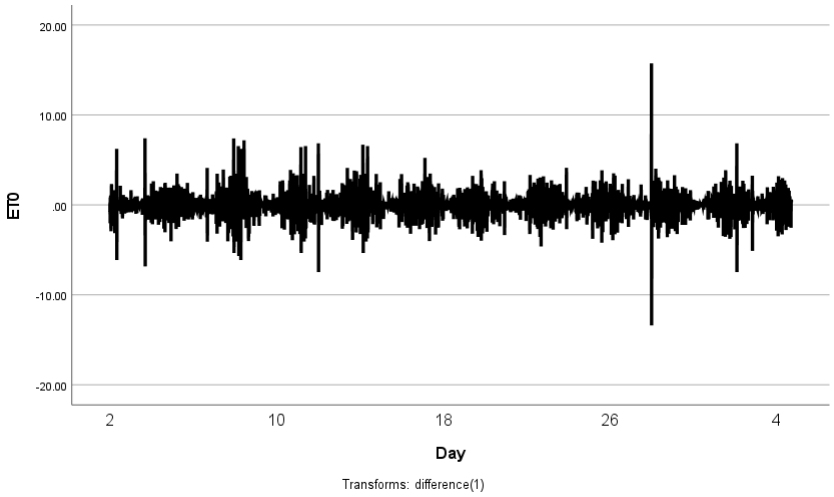


Figure 4.15: First difference time series plot of water evaporation rate (Leeukraal DeHoop Dam).

Figure 4.16 presents that the first differenced time series plot resulted to be stationary after differencing. The water evaporation rate trend shows an accurate pattern, where the mean and the variance of the water evaporation rate time series is constant, hence there is no need for second differencing (Adhikari and Agrawal, 2013; Shumway et al., 2000).

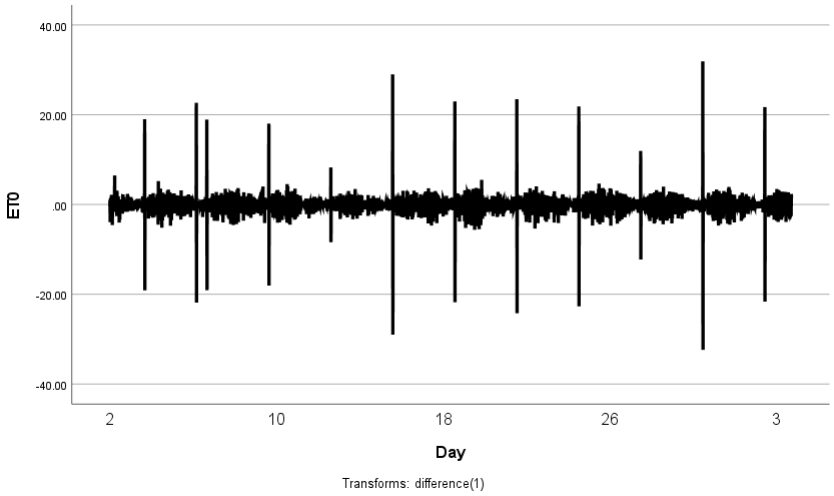


Figure 4.16: First difference time series plot of water evaporation rate (Luphephe Dam).

Figure 4.17 presents the first differenced Autocorrelation Function plot of water evaporation rate in Mokolo dam where by from lag 1 to lag 16 the plot is stationary. The plot shows that at lag 1 and 2 the peaks are significant whereas the rest of the peaks are insignificant. The ACF of the series declared that the suggested model can be MA(1) and MA(2) (Nau, 2014).

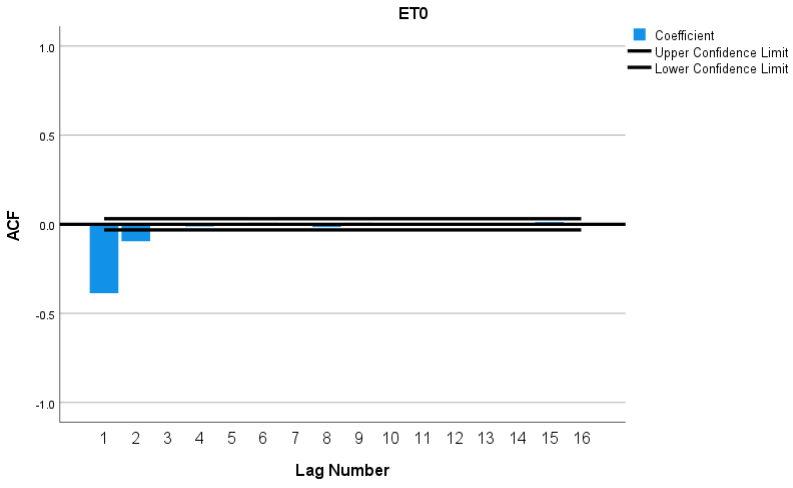


Figure 4.17: First difference autocorrelation function of water evaporation rate (Mokolo Dam)

Figure 4.18 presents the first differenced partial autocorrelation function plot of water evaporation rate in Mokolo dam where by from lag 1 to lag 16 the plot is stationary. The plot shows that from lag 1 to lag 14 the peaks are significant where as the rest of th of the peaks are insignificant. The ACF of the series declared that the suggested model can be MA (1), MA (2) and MA (3) (Nau, 2014).

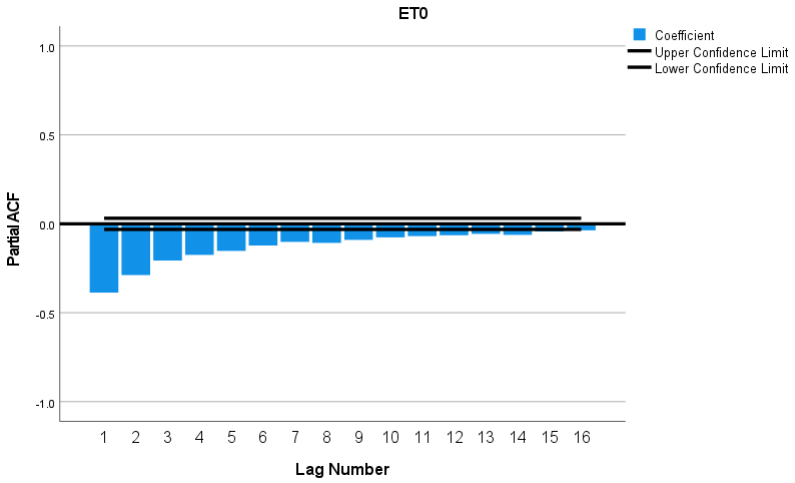


Figure 4.18: First difference Partial Autocorrelation Function of water evaporation rate (Mokolo Dam).

Figure 4.19 presents the first differenced autocorrelation function plot of water evaporation rate in Ga-Rantho dam where by from lag 1 to lag 16 the plot is stationary. The plot shows that at lag 1 and 2 the peaks are significant whereas the rest of the peaks are insignificant. The ACF of the series declared that the suggested model can be MA(1) and MA(2) (Nau, 2014).

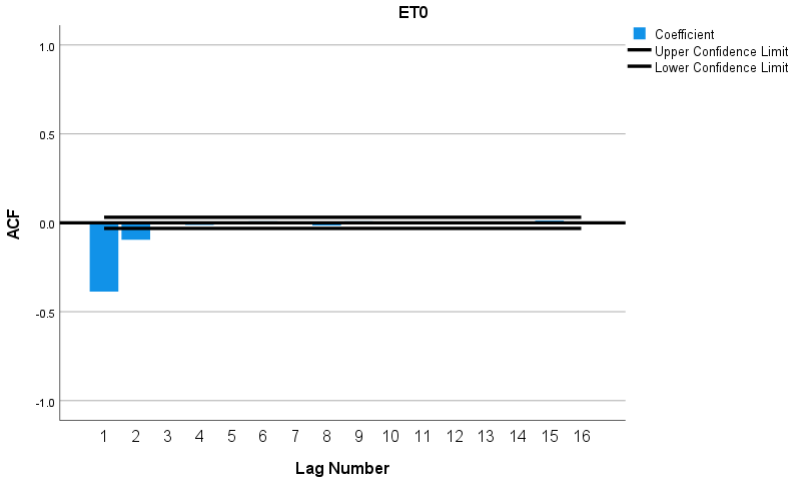


Figure 4.19: First difference Autocorrelation Function of water evaporation rate (Ga-Rantho Dam).

Figure 4.20 presents the first differenced partial autocorrelation function plot of water evaporation rate in Mokolo dam from lag 1 to lag 16 is stationary. The plot shows that at from 1 to lag 14 the peaks are significant whereas the rest of the peaks are insignificant. The ACF of the series declared that the suggested model can be MA (1), MA (2) and MA (3) (Nau, 2014).

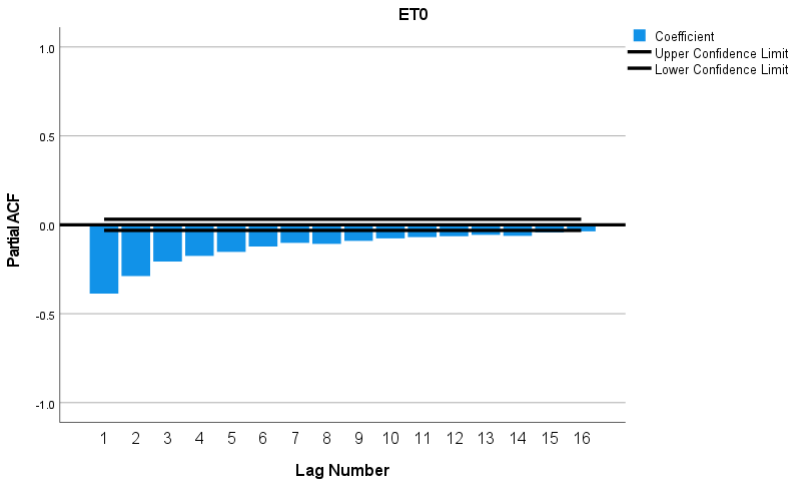


Figure 4.20: First difference Partial Autocorrelation Function of water evaporation rate (Ga-Rantho Dam).

Figure 4.21 presents the first differenced autocorrelation function plot of water evaporation rate in Leeukraal DeHoop where by from lag 1 to lag 16 the plot is stationary. The plot shows that at lag 1 and 2 the peaks are significant whereas the rest of the of the peaks are insignificant. The ACF of the series showed that the suggested model can be MA(1), MA(2) and MA(3) (Nau, 2014).

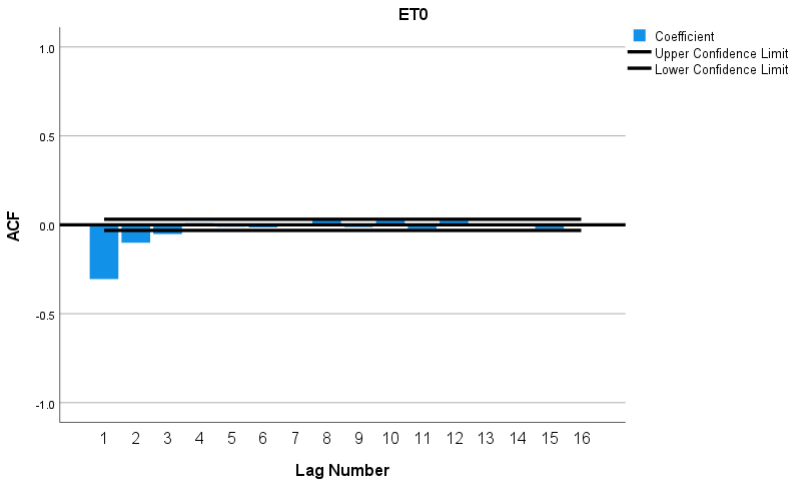


Figure 4.21: First difference Autocorrelation Function of water evaporation rate (Leeukraal DeHoop Dam).

Figure 4.22 presents the first differenced partial autocorrelation function plot of water evaporation rate in Leeukraal DeHoop dam where

Figure 23 presents the first differenced autocorrelation function plot of water evaporation rate in Luphephe dam where by from lag 1 to lag 16 the plot is stationary. The plot shows that at lag 1 and 2 the peaks are significant whereas the rest of the peaks are insignificant. The ACF of the series revealed that the suggested model can be MA (1) and MA (2) (Nau, 2014).

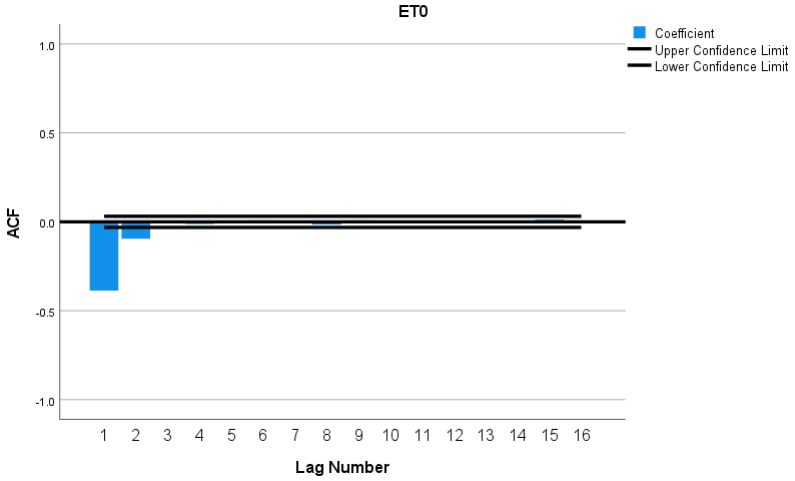


Figure 4.23: First difference Autocorrelation Function of water evaporation rate (Luphephe Dam).

Figure 4.24 presents the first differenced partial autocorrelation function plot of water evaporation rate in Luphephe dam where by from lag 1 to lag 16 the plot is stationary. The plot shows that from lag 1 to lag 14 the peaks are significant whereas the rest of the peaks are insignificant. The ACF of the series revealed that the suggested model can be MA(1), MA(2) and MA (3) (Nau, 2014).

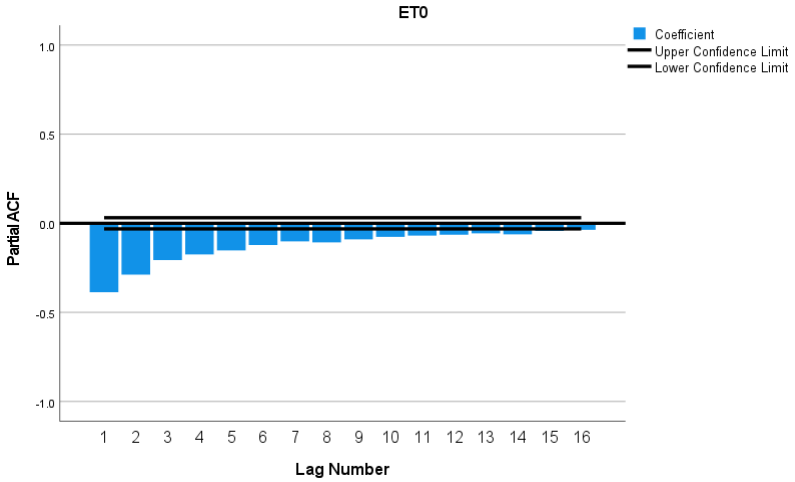


Figure 4.24: First difference Partial Autocorrelation Function of water evaporation rate (Luphephe Dam).

4.2.3 Testing for stationary Augmented Dikey Fuller test of water evaporation rate from the selected dams

Table 4.2: Agumented Dicky Fuller test for water evaporation rate Mokolo dam.

	t-statistic	$Prob.^*$
Agumented Dicky-Fuller test statistics	-2.923	$2e^{-16}$
Test critical values		
1% level	-3.96	
5% level	-3.41	
10% level	-3.12	

Table 4.2 presents the Agumented Dickey fuller (ADF) test for water evaporation rate in Mokolo dam. The water evaporation rate time series have non-stationary process since ADF test statistic of -2.923 is greather than the critical value at 1%, 5% and 10% (Dwivedi et al., 2017; Fathian, 2019).

Table 4.3: Agumented Dicky Fuller test for water evaporation rate Ga-Ranth dam.

	t-statistic	$Prob.^*$
Agumented Dicky-Fuller test statistics	-1.9057	0.609
Test critical values		
1% level	-3.96	
5% level	-3.41	
10% level	-3.12	

Table 4.3 presents the Agumented Dickey fuller (ADF) test for water evaporation rate in Ga-Rantho dam. The water evaporation rate time series have non-stationary process since, ADF test statistic of -1.9057 is greather than the critical value at 1%, 5% and 10% (Dwivedi et al., 2017; Fathian, 2019).

Table 4.4: Agumented Dicky Fuller test for water evaporation rate Leeukraal DeHoop dam.

	t-statistic	$Prob.^*$
Agumented Dicky-Fuller test statistics	-1.6941	$2e^{-16}$
Test critical values		
1% level	-3.96	
5% level	-3.41	
10% level	-3.12	

Table 4.4 presents the Agumented Dickey fuller (ADF) test for water evaporation rate in Leeukraal DeHoop dam. The water evaporation rate time series have non-stationary process since, ADF test statistic of -1.6941 is greater than the critical value at 1%, 5% and 10% (Dwivedi et al., 2017; Fathian, 2019).

Table 4.5: Agumented Dicky Fuller test for water evaporation rate Luphephe dam.

	t-statistic	$Prob.^*$
Agumented Dicky-Fuller test statistics	-3.115	$2e^{-16}$
Test critical values		
1% level	-3.96	
5% level	-3.41	
10% level	-3.12	

Table 4.5 presents the Agumented Dickey fuller (ADF) test for water evaporation rate in Luphephe dam. The water evaporation rate time series have non-stationary process since, ADF test statistic of -3.115 is greater than the critical value at 1%, 5% and 10% (Dwivedi et al., 2017; Fathian, 2019).

Table 4.6: First differenced Agumented Dicky Fuller test for water evaporation rate Mokolo dam.

	t-statistic	$Prob.^*$
Agumented Dicky-Fuller test statistics	-11.855	$1.28e^{-12}$
Test critical values		
1% level	-3.96	
5% level	-3.41	
10% level	-3.12	

Table 4.6 presents the first differenced Agumented Dickey fuller (ADF) test for water evaporation rate in Mokolo dam. The water evaporation rate time series have a stationary process since, ADF test statistic of -11.855 is less than the critical value at 1%, 5% and 10% (Dwivedi et al., 2017; Fathian, 2019).

Table 4.7: First differenced Agumented Dicky Fuller test for water evaporation rate Ga-Rantho dam.

	t-statistic	$Prob.^*$
Agumented Dicky-Fuller test statistics	-7.118	$2e^{-16}$
Test critical values		
1% level	-3.96	
5% level	-3.41	
10% level	-3.12	

Table 4.7 presents the first differenced Agumented Dickey fuller (ADF) test for water evaporation rate in Ga-Rantho dam. The water evaporation rate time series have stationary process since, ADF test statistic of -7.118 is greater than the critical value at 1%, 5% and 10% (Dwivedi et al., 2017; Fathian, 2019).

Table 4.8: First differenced Agumented Dicky Fuller test for water evaporation rate Leeukraal DeHoop dam.

	t-statistic	$Prob.^*$
Agumented Dicky-Fuller test statistics	-14.392	$2e^{-16}$
Test critical values		
1% level	-3.96	
5% level	-3.41	
10% level	-3.12	

Table 4.8 presents the first differenced Agumented Dickey fuller (ADF) test for water evaporation rate in Mokolo dam. The water evaporation rate time series have a stationary process since, ADF test statistic of -14.398 is greater than the critical value at 1%, 5% and 10% (Dwivedi et al., 2017; Fathian, 2019).

Table 4.9: First differenced Agumented Dicky Fuller test for water evaporation rate Luphephe dam.

	t-statistic	$Prob.^*$
Agumented Dicky-Fuller test statistics	-7.118	$1.\ 28e^{-12}$
Test critical values		
1% level	-3.96	
5% level	-3.41	
10% level	-3.12	

Table 4.9 presents the first differenced Agumented Dickey fuller (ADF) test for water evaporation rate in Mokolo dam. The water evaporation rate time series have a stationary process since ADF test statistic of -7.118 is greater than the critical value at 1%, 5% and 10% (Dwivedi et al., 2017; Fathian, 2019).

4.2.4 Model identification

In this section, the models are to be identified and the best model will be drawn from the selected models. The models identified are suggested from Figure 17 to Figure 24 of the first non-seasonal ACF and PACF. The models that are suggested from figure 17 to figure 24 are AR (1), AR (2), AR (3) AND MA(1), MA (2), MA (3).

4.2.5 Model estimation

Table 4.10: ARIMA(p,d,q) model summary of water evaporation rate Mokolo dam.

Models	ARIMA(p,d,q)
1	ARIMA(1,1,1)
2	ARIMA(1,1,2)
3	ARIMA(2,1,3)

Table 4.11: ARIMA(p,d,q) model summary of water evaporation rate Ga-Rantho dam.

Models	ARIMA(p,d,q)
1	ARIMA(1,1,2)
2	ARIMA(1,1,3)
3	ARIMA(2,1,2)

Table 4.12: ARIMA(p,d,q) model summary of water evaporation rate Lueekraal DeHoop dam.

Models	ARIMA(p,d,q)
1	ARIMA(1,1,1)
2	ARIMA(1,1,3)
3	ARIMA(2,1,1)

Table 4.13: ARIMA(p,d,q) model summary of water evaporation rate Luphephe dam.

Models	ARIMA(p,d,q)
1	ARIMA(1,1,3)
2	ARIMA(2,1,2)
3	ARIMA(2,1,3)

Table 4.10, table 4.11, table 4.12 and table 4.13 present the summarised selected ARIMA models among other models using the principle of parsimony from the selected dams.

Table 4.14: Model fit for ARIM(1,1,1) of water evaporation rate Mokolo dam.

Fit statistic	Water evaporation rate
Stationary R-squared	0.380
R-squared	0.233
RMSE	1.674
MAPE	22.941
MAE	0.778
Normalised BIC	1.039

Table 4.15: Model statistic for ARIMA (1,1,1) model for water evaporation Mokolo dam.

Model	LJUNG-Box		
	Statistic	DF	p-value
ARIMA	13.315	16	0.65

Table 4.14 and 4.15 shows the model fit and model statistics for ARIMA (1,1,1). The value of RMSE, MAPE, MAE and BIC are presented where the value of RMSE does not shows a fit for the model, hence the model cannot relatively predict the data accurately. The MAPE is equal to 22.941, RMSE is equal to 1.674, MAE is equal to 0.778 and BIC is equal to 1.039. The value of BIC cannot be compared with anything since there is no BIC value to compared it with from the model. The Ljung-Box p-value is equal to 0.65 which indicates that it is insignificant at 5% level of significant (Dabral et al., 2017; Issaka, 2015).

Table 4.16: Parameter estimation for ARIMA(1,1,1) model for water evaporation Mokolo dam.

Model	AR(1)	MA(1)
Estimate	0.143	0.918
SE	0.018	0.007
t	8.165	130.315
P-value	0.000	0.000

Table 4.16 presents the parameter estimation for ARIMA (1,1,1) model, the two models AR and MA are significant. The AR (1) and MA (1) model are significant since the p-value is less than 5% with parameter estimate of AR (1) = 0.143 and MA (1) = 0.918. This results indicate that water evaporation rate at Mokolo dam increases.

Table 4.17: Model fit for ARIM(1,1,2) of water evaporation rate Ga-Ranth dam.

Fit statistic	Water evaporation rate
Stationary R-squared	0.793
R-squared	0.893
RMSE	0.543
MAPE	20.283
MAE	1.733
Normalised BIC	0.139

Table 4.18: Model statistic for ARIMA (1,1,2) model for water evaporation Ga-Rantho dam.

Model	LJUNG-Box		
	Statistic	DF	p-value
ARIMA	1.930	15	0.010

Table 4.17 and 4.18 shows the model fit and model statistics for ARIMA (1,1,2). The value of RMSE, MAPE, MAE and BIC are presented where the value of RMSE shows a fit for the model, hence the model can relatively predict the data accurately. The MAPE is equal to 20.283, RMSE is equal to 0.543, MAE is equal to 1.733 and BIC is equal to 0.139. The value of BIC cannot be compared with anything since there is no BIC value to compared it with from the model. The Ljung-Box p-value is equal to 0.010 which indicates that it is significant at 5% level of significant(Dabral et al., 2017; Issaka, 2015).

Table 4.19: Parameter estimation for ARIMA(1,1,2) model for water evaporation Ga-Rantho dam.

Model	AR (1)	MA (2)
Estimate	0.319	-0.178
SE	0.129	0.001
t	2.474	-1.480
P-value	0.013	0.139

Table 4.19 presents the parameter estimation for ARIMA (1,1,2) model, the two models AR and MA are significant. The AR (1) model is significant is less than 5% and MA (2) model is insignificant since the p-value is more than 5% with parameter estimate of AR (1)= 0.319 and MA (2)= -0.178. This results indicate that water evaporation rate at Ga-Rantho dam inconclusive whether it decreases or increases.

Table 4.20: Model fit for ARIM(1,1,1) of water evaporation rate Leeukraal DeHoop dam.

Fit statistic	Water evaporation rate
Stationary R-squared	0.193
R-squared	0.668
RMSE	0.385
MAPE	18.027
MAE	0.706
Normalised BIC	0.171

Table 4.21: Model statistic for ARIMA (1,1,1) model for water evaporation Leeukraaal DeHoop dam.

Model	LJUNG-Box		
	Statistic	DF	p-value
ARIMA	17.394	16	0.361

Table 4.20 and 4.21 shows the model fit and model statistics for ARIMA (1,1,1). The value of RMSE, MAPE, MAE and BIC are presented where the value of RMSE does not shows a fit for the model, hence the model cannot relatively predict the data accurately. The MAPE is equal to 24.039, RMSE is equal to 1.085, MAE is equal to 0.706 and BIC is equal to 0.171. The value of BIC cannot be compared with anything since there is no BIC value to compared it with from the model. The Ljung-Box p-value is equal to 0.361 which indicates that it is insignificant at 5% level of significant (Dabral et al., 2017; Issaka, 2015).

Table 4.22: Parameter estimation for ARIMA(1,1,1) model for water evaporation Leeukraal DeHoop dam.

Model	AR (1)	MA (1)
Estimate	0.327	0.790
SE	0.024	0.001
t	13.743	51.020
P-value	0.000	0.000

Table 4.22 presents the parameter estimation for ARIMA (1,1,1) model, the two models AR and MA are significant. The AR (1) model is significant is less than 5% and MA (1) model is significant since the p-value is less than 5% with parameter estimate of AR (1)= 0.327 and MA (1)= 0.790. This results indicate that water evaporation rate at Leeukraal DeHoop dam increases.

Table 4.23: Model fit for ARIM(1,1,3) of water evaporation rate Luphephe dam.

Fit statistic	Water evaporation rate
Stationary R-squared	0.350
R-squared	0.186
RMSE	1.728
MAPE	55.961
MAE	1.016
Normalised BIC	1.107

Table 4.24: Model statistic for ARIMA (1,1,3) model for water evaporation Luphephe dam.

Model	LJUNG-Box		
	Statistic	DF	p-value
ARIMA	10.792	14	0.702

Table 4.23 and 4.24 shows the model fit and model statistics for ARIMA (1,1,3). The value of RMSE, MAPE, MAE and BIC are presented where the value of RMSE does not shows a fit for the model, hence the model cannot relatively predict the data accurately. The MAPE is equal to 55.961, RMSE is equal to 1.728, MAE is equal to 1.016 and BIC is equal to 1.107. The value of BIC cannot be compared with anything since there is no BIC value to compared it with from the model. The Ljung-Box p-value is equal to 0.702 which indicates that it is insignificant at 5% level of significant (Dabral et al., 2017; Issaka, 2015).

Table 4.25: Parameter estimation for ARIMA(1,1,3) model for water evaporation Luphephe dam.

Model	AR (1)	MA (3)
Estimate	0.938	-0.159
SE	0.023	0.019
t	40.936	-8.326
P-value	0.000	0.000

Table 4.25 presents the parameter estimation for ARIMA (1,1,3) model, the two models AR and MA are significant. The AR (1) model is significant is less than 5% and MA (3) model is significant since the p-value is less than 5% with parameter estimate of AR (1)= 0.938 and MA (3)= -0.159. This results indicate that water evaporation rate at Luphephe dam increases.

Table 4.26: Model fit for ARIM(1,1,2) of water evaporation rate Mokolo dam.

Stationary R-squared	0.381
R-squared	0.734
RMSE	0.524
MAPE	19.023
MAE	0.730
Normalised BIC	0.872

Table 4.27: Model statistic for ARIMA (1,1,2) model for water evaporation Mokolo dam.

Model	LJUNG-Box		
	Statistic	DF	p-value
ARIMA	11.654	15	0.705

Table 4.26 and 4.27 shows the model fit and model statistics for ARIMA (1,1,2). The value of RMSE, MAPE, MAE and BIC are presented where the value of RMSE shows a fit for the model, hence the model cannot relatively predict the data accurately. The MAPE is equal to 19.023, RMSE is equal to 0.524, MAE is equal to 0.730 and BIC is equal to 0.872. The value of BIC cannot be compared with anything since there is no BIC value to compared it with from the model. The Ljung-Box p-value is equal to 0.705 which indicates that it is insignificant at 5% level of significant (Dabral et al., 2017; Issaka, 2015)

Table 4.28: Parameter estimation for ARIMA(1,1,2) model for water evaporation Mokolo dam.

Model	AR(1)	MA(2)
Estimate	0.469	-0.298
SE	0.099	0.095
t	4.730	-3.149
P-value	0.000	0.002

Table 4.28 presents the parameter estimation for ARIMA (1,1,2) model, the two models AR and MA are significant. The AR (1) model is significant is less than 5% and MA (2) model is significant since the p-value is less than 5% with parameter estimate of AR (1)= 0.469 and MA (2)= -0.298. This results indicate that water evaporation rate at Mokolo dam increases.

Table 4.29: Model fit for ARIM(1,1,3) of water evaporation rate Ga-Ranth dam.

Fit statistic	Water evaporation rate
Stationary R-squared	0.393
R-squared	0.293
RMSE	1.673
MAPE	25.280
MAE	0.733
Normalised BIC	1.041

Table 4.30: Model statistic for ARIMA (1,1,3) model for water evaporation Ga-Rantho dam.

Model	LJUNG-Box		
	Statistic	DF	p-value
ARIMA	1.896	14	1.000

Table 4.29 and 4.30 shows the model fit and model statistics for ARIMA (1,1,3). The value of RMSE, MAPE, MAE and BIC are presented where the value of RMSE does not shows a fit for the model, hence the model cannot relatively predict the data accurately. The MAPE is equal to 25.280, RMSE is equal to 1.673, MAE is equal to 0.733 and BIC is equal to 1.041. The value of BIC cannot be compared with anything since there is no BIC value to compared it with from the model. The Ljung-Box p-value is equal to 1.000 which indicates that it is insignificant at 5% level of significant (Dabral et al., 2017; Issaka, 2015).

Table 4.31: Parameter estimation for ARIMA(1,1,3) model for water evaporation Ga-Rantho dam.

Model	AR (1)	MA (3)
Estimate	0.263	0.007
SE	0.419	0.050
t	0.626	0.149
P-value	0.531	0.882

Table 4.31 presents the parameter estimation for ARIMA (1,1,3) model, the two models AR and MA are significant. The AR (1) model is insignificant is more than 5% and MA (3) model is insignificant since the p-value is more than 5% with parameter estimate of AR (1)= 0.263 and MA (3)= 0.007. This results indicates that water evaporation rate at Ga-Rantho dam decreases.

Table 4.32: Model fit for ARIM(1,1,3) of water evaporation rate Leeukraal DeHoop dam.

Fit statistic	Water evaporation rate
Stationary R-squared	0.194
R-squared	0.668
RMSE	1.085
MAPE	24.014
MAE	0.706
Normalised BIC	0.175

Table 4.33: Model statistic for ARIMA (1,1,3) model for water evaporation Leeukraaal DeHoop dam.

Model	LJUNG-Box		
	Statistic	DF	p-value
ARIMA	15.273	14	0.360

Table 4.32 and 4.33 shows the model fit and model statistics for ARIMA (1,1,3). The value of RMSE, MAPE, MAE and BIC are presented where the value of RMSE does not shows a fit for the model, hence the model cannot relatively predict the data accurately. The MAPE is equal to 24.014, RMSE is equal to 1.085, MAE is equal to 0.706 and BIC is equal to 0.175. The value of BIC cannot be compared with anything since there is no BIC value to compared it with from the model. The Ljung-Box p-value is equal to 0.360 which indicate that it is insignificant at 5% level of significant (Dabral et al., 2017; Issaka, 2015).

Table 4.34: Parameter estimation for ARIMA(1,1,3) model for water evaporation Leeukraal DeHoop dam.

Model	AR(1)	MA(3)
Estimate	-0.031	0.072
SE	0.237	0.046
t	-0.131	1.552
P-value	0.896	0.121

Table 4.34 presents the parameter estimation for ARIMA (1,1,3) model, the two models AR and MA are significant. The AR (1) model is insignificant is more than 5% and MA (3) model is insignificant since the p-value is more than 5% with parameter estimate of AR (1)= -0.031 and MA (3)= 0.072. This results indicate that water evaporation rate at Leeukraal DeHoop dam decreases.

Table 4.35: Model statistic for ARIMA (2,1,2) model for water evaporation Luphephe dam.

Fit statistic	Water evaporation rate
Stationary R-squared	0.348
R-squared	0.184
RMSE	1.731
MAPE	55.843
MAE	1.010
Normalised BIC	1.109

Table 4.36: Model statistic for ARIMA (2,1,3) model for water evaporation Mokolo dam.

Model	LJUNG-Box		
	Statistic	DF	p-value
ARIMA	5.122	14	0.984

Table 4.35 and 4.36 shows the model fit and model statistics for ARIMA (2,1,2). The value of RMSE, MAPE, MAE and BIC are presented where the value of RMSE does not shows a fit for the model, hence the model cannot relatively predict the data accurately. The MAPE is equal to 55.843, RMSE is equal to 1.731, MAE is equal to 1.010 and BIC is equal to 1.109. The value of BIC cannot be compared with anything since there is no BIC value to compared it with from the model. The Ljung-Box p-value is equal to 0.984 this indicates that it is insignificant at 5% level of significant (Dabral et al., 2017; Issaka, 2015).

Table 4.37: Parameter estimation for ARIMA(2,1,2) model for water evaporation Luphephe dam.

Model	AR (2)	MA (2)
Estimate	0.204	0.863
SE	0.059	0.207
t	3.480	4.168
P-value	0.001	0.000

Table 4.37 presents the parameter estimation for ARIMA (2,1,2) model, the two models AR and MA are significant. The AR (2) model is significant is less than 5% and MA (2) model is significant since the p-value is less than 5% with parameter estimate of AR (2)= 0.204 and MA (2)= 0.863. This results indicate that water evaporation rate at Luphephe dam increases.

Table 4.38: Model fit for ARIM(2,1,3) of water evaporation rate Mokolo dam.

Fit statistic	Water evaporation rate
Stationary R-squared	0.382
R-squared	0.236
RMSE	1.672
MAPE	23.019
MAE	0.782
Normalised BIC	1.042

Table 4.39: Model statistic for ARIMA (2,1,3) model for water evaporation Mokolo dam.

Model	LJUNG-Box		
	Statistic	DF	p-value
ARIMA	5.094	13	0.973

Table 4.38 and 4.39 shows the model fit and model statistics for ARIMA (2,1,3). The value of RMSE, MAPE, MAE and BIC are presented where the value of RMSE does not shows a fit for the model, hence the model cannot relatively predict the data accurately. The MAPE is equal to 23.019, RMSE is equal to 1.672, MAE is equal to 0.782 and BIC is equal to 1.042. The value of BIC cannot be compared with anything since there is no BIC value to compared it with from the model. The Ljung-Box p-value is equal to 0.973 which indicates that it is insignificant at 5% level of significant (Dabral et al., 2017; Issaka, 2015).

Table 4.40: Parameter estimation for ARIMA (2,1,3) model for water evaporation Mokolo dam.

Model	AR(2)	MA(3)
Estimate	0.127	-0.172
SE	0.130	0.096
t	0.974	-1.802
P-value	0.330	0.072

Table 4.40 presents the parameter estimation for ARIMA (2,1,2) model, the two models AR and MA are significant. The AR (2) model is insignificant is more than 5% and MA (3) model is insignificant since the p-value is more than 5% with parameter estimate of AR (2)= 0.127 and MA (3)= -0.172. This results indicate that water evaporation rate at Mokolo dam decreases.

Table 4.41: Model fit for ARIM(2,1,2) of water evaporation rate Ga-Ranth dam.

Fit statistic	Water evaporation rate
Stationary R-squared	0.393
R-squared	0.293
RMSE	1.673
MAPE	25.282
MAE	0.733
Normalised BIC	1.041

Table 4.42: Model statistic for ARIMA (2,1,2) model for water evaporation Ga-Rantho dam.

Model	LJUNG-Box		
	Statistic	DF	p-value
ARIMA	1.898	14	1.000

Table 4.41 and 4.42 shows the model fit and model statistics for ARIMA (2,1,2). The value of RMSE, MAPE, MAE and BIC are presented where the value of RMSE does not shows a fit for the model, hence the model cannot relatively predict the data accurately. The MAPE is equal to 25.282, RMSE is equal to 1.673, MAE is equal to 0.733 and BIC is equal to 1.041. The value of BIC cannot be compared with anything since there is no BIC value to compared it with from the model. The Ljung-Box p-value is equal to 1.000 which indicate that it is insignificant at 5% level of significant (Dabral et al., 2017; Issaka, 2015).

Table 4.43: Parameter estimation for ARIMA (2,1,2) model for water evaporation Ga-Rantho dam.

Model	AR (2)	MA (2)
Estimate	0.014	-0.079
SE	0.091	0.653
t	0.158	-0.121
P-value	0.874	0.904

Table 4.43 presents the parameter estimation for ARIMA (2,1,2) model, the two models AR and MA are significant. The AR (2) model is insignificant is more than 5% and MA (2) model is insignificant since the p-value is more than 5% with parameter estimate of AR (2)= 0.014 and MA (2)= -0.079. This results indicate that water evaporation rate at Ga-Rantho dam decreases.

Table 4.44: Model fit for ARIM(2,1,1) of water evaporation rate Leeukraal DeHoop dam.

Fit statistic	Water evaporation rate
Stationary R-squared	0.193
R-squared	0.668
RMSE	1.085
MAPE	24.039
MAE	0.506
Normalised BIC	0.173

Table 4.45: Model statistic for ARIMA (2,1,1) model for water evaporation Leeukraaal DeHoop dam.

Model	LJUNG-Box		
	Statistic	DF	p-value
ARIMA	17.216	15	0.036

Table 4.44 and 4.45 shows the model fit and model statistics for ARIMA (2,1,1). The value of RMSE, MAPE, MAE and BIC are presented where the value of RMSE shows a fit for the model, hence the model cannot relatively predict the data accurately. The MAPE is equal to 18.027, RMSE is equal to 0.385, MAE is equal to 0.506 and BIC is equal to 0.173. The value of BIC cannot be compared with anything since there is no BIC value to compared it with from the model. The Ljung-Box p-value is equal to 0.036 which indicates that it is significant at 5% level of significant (Dabral et al., 2017; Issaka, 2015).

Table 4.46: Parameter estimation for ARIMA (2,1,1) model for water evaporation Leeukraal DeHoop dam.

Model	AR (2)	MA (1)
Estimate	-0.011	0.784
SE	0.020	0.020
t	-0..562	39.563
P-value	0.574	0.788

Table 4.46 presents the parameter estimation for ARIMA (2,1,1) model, the two models AR and MA are significant. The AR (2) model is insignificant is less than 5% and MA (1) model is insignificant since the p-value is more than 5% with parameter estimate of AR (2)= -0.011 and MA (1)= 0.784. This results indicate that water evaporation rate at Leeukraal DeHoop dam decreases.

Table 4.47: Model fit for ARIM(2,1,3) of water evaporation rate Luphephe dam.

Fit statistic	Water evaporation rate
Stationary R-squared	0.350
R-squared	0.886
RMSE	0.479
MAPE	50.966
MAE	0.416
Normalised BIC	0.113

Table 4.48: Model statistic for ARIMA (2,1,3) model for water evaporation Luphephe dam.

Model	LJUNG-Box		
	Statistic	DF	p-value
ARIMA	11.203	13	0.054

Table 4.47 and 4.48 shows the model fit and model statistics for ARIMA (2,1,3). The value of RMSE, MAPE, MAE and BIC are presented where the value of RMSE shows a fit for the model, hence the model cannot relatively predict the data accurately. The MAPE is equal to 55.966, RMSE is equal to 0.479, MAE is equal to 0.416 and BIC is equal to 0.117. The value of BIC cannot be compared with anything since there is no BIC value to compared it with from the model. The Ljung-Box p-value is equal to 0.054 which indicates that it insignificant at 5% level of significant (Dabral et al., 2017; Issaka, 2015).

Table 4.49: Parameter estimation for ARIMA (2,1,3) model for water evaporation Luphephe dam.

Model	AR (2)	MA (3)
Estimate	0.016	-0.167
SE	0.025	0.030
t	0.648	-5.659
P-value	0.517	0.000

Table 4.49 presents the parameter estimation for ARIMA (2,1,3) model, the two models AR and MA are significant. The AR (2) model is insignificant is more than 5% and MA (3) model is significant since the p-value is less than 5% with parameter estimate of AR (2)= 0.016 and MA (3)= -0.167. This results indicate that water evaporation rate at Luphephe dam inconclusive whether it decreases or increases.

4.2.6 VAR model Coefficients and Correlation residuals relationship between water evaporation and explanatory variables from the selected dams

Table 4.50: VAR model Coefficients Mokolo dam.

Coefficients	Estimate	Std.Error	t value	Pr$\|>t\|$
Evaporation	0.233883	0.005329	43.89	$2(e^{-16})$
Temperature	0.117747	0.004558	25.84	$2.2(e^{-16})$
Rain	0.001131	0.00168	0.67	0.503

Table 4.50 presents coefficients with estimation of evaporation (0.23322), temperature (0.11774) and rain (0.00113). The p-values of evaporation and temperature are significant, the p-value is less than 5% level of significant. Rain is insignificant for Mokolo dam with p-value grater than 5% level of significant (Issaka, 2015).

Table 4.51: VAR correlation residual relationship of water evaporation and (temperature and rainfall) for Mokolo dam.

Correlation of residuals				
	EVAPO	MAX TEMP	MIN TEMP	RAIN
EVAPORATION	1.0000000	0.395562	0.011703	0.0005259
MAX TEMP	0.3955617	1.000000	0.087236	0.0023128
MIN TEMP	0.0117027	0.087236	1.000000	-0.0066775
RAIN	0.0005259	0.002313	-0.006677	1.0000000

Table 4.51 presents the relationship between evaporation with temperature and rainfall. The value of evaporation and minimum temperature (0.0117027), evaporation and maximum temperature (0.3955617), and evaporation and rainfall (0.0005259) indicates a weak positive relationship (Issaka, 2015).

Table 4.52: VAR model Coefficients Ga-Rantho dam.

Coefficients	Estimate	Std.Error	t value	Pr(t)
Evaporation	0.324215	0.005739	38.6589	$2(e^{-16})$
Temperature	0.155789	0.004813	32.37	$2(e^{-16})$
Rain	0.004489	0.004387	1.527	0.07523

Table 4.52 presents coefficients with estimation of evaporation (0.32422), temperature (0.15573) and rain (0.004489). The p-values of evaporation and temperature are significant with ($p < 0.05$). Rain variable is insignificant for Ga-Rantho dam with ($p > 0.05$) (Issaka, 2015).

Table 4.53: VAR correlation residual relationship of water evaporation and (temperature and rainfall) for Ga-Rantho dam.

Correlation of residuals				
	EVAPO	MAX TEMP	MIN TEMP	RAIN
EVAPO	1.00000	0.308436	-0.01711	-0.11641
MAX TEMP	0.30836	1.0000	0.14174	-0.11009
MIN TEMP	-0.01711	0.1417	1.00000	-0.01023
RAIN	-0.11641	-0.1101	-0.01023	1.00000

Table 4.53 presents the relationship between evaporation with temperature and rainfall. The value of evaporation and maximum temperature (0.30836) indicates a weak positive relationship, evaporation and minimum temperature (-0.01711) indicates a weak negative relationship, and evaporation and rainfall (-0.11641) indicates a weak negative relationship.

Table 4.54: VAR model Coefficients Leeukraal DeHoop dam.

Coefficients	Estimate	Std.Error	t value	Pr(t)
Evaporation	0.278691	0.004753	47.48	$2(e^{-16})$
Temperature	0.162389	0.004813	32.37	$2.2(e^{-16})$
Rain	0.008719	0.005419	1.652	0.08655

Table 4.54 presents coefficients with estimation of evaporation (0.27869), temperature (0.162389) and rain (0.00872). The p-values of evaporation and temperature are significant with ($p < 0.05$). Only the rain variable is insignificant with ($p < 0.05$) (Issaka, 2015).

Table 4.55: VAR correlation residual relationship of water evaporation and (temperature and rainfall) for Leeukraal DeHoop dam.

Correlation of residuals				
	EVAPO	MAX TEMP	MIN TEMP	RAIN
EVAPO	1.00000	0.53557	-0.05324	-0.15583
MAX TEMP	0.53557	1.00000	0.14532	-0.06993
MIN TEMP	-0.05324	0.14532	1.00000	0.07376
RAIN	-0.15583	-0.06993	0.07376	1.00000

Table 4.55 present the relationship between evaporation with temperature and rainfall. The value of evaporation and maximum temperature (0.53557) indicates a positive relationship, evaporation and minimum temperature (-0.05324) indicates a weak negative relationship, and evaporation and rainfall (-0.15583) indicates a weak negative relationship (Issaka, 2015).

Table 4.56: VAR model Coefficients Luphephe dam.

Coefficients	Estimate	Std.Error	t value	Pr(t)
Evaporation	0.238215	0.004454	53.48	$2(e^{-16})$
Temperature	0.155789	0.004813	32.37	$2.2(e^{-16})$
Rain	0.007591	0.004569	1.661	0.0967

Table 4.56 presents coefficients with estimation of evaporation (0.23821), temperature (0.15579) and rain (0.00457). The p-values of evaporation and temperature are significant with ($p < 0.05$). The p-value for rain is insignificant with ($p > 0.05$) (Issaka, 2015).

Table 4.57: VAR correlation residual relationship of water evaporation and (temperature and rainfall) for Luphephe dam.

Correlation of residuals				
	EVAPO	MAX TEMP	MIN TEMP	RAIN
EVAPORATION	1.00000	0.99951	0.04727	0.01289
MAX TEMP	0.99951	1.00000	0.66025	0.10619
MIN TEMP	0.04727	0.66025	1.00000	-0.24497
RAIN	0.01289	0.10619	-0.24497	1.00000

Table 4.57 presents the relationship between evaporation with temperature and rainfall. The value of evaporation and maximum temperature (0.99951) indicates a strong positive relationship, evaporation and minimum temperature (0.04727) indicates a weak positive relationship, and evaporation and rainfall (0.01289) indicates weak positive relationship (Issaka, 2015).

4.2.7 ARCH and model of water evaporation from the selected dams.

Table 4.58: ARCH model of water evaporation for Mokolo dam.

Weighted ARCH LM Tests				
	Statistic	Shape	Scale	P-Value
ARCH Lag[3]	0.00599	0.500	2.000	0.9383
ARCH Lag[5]	0.01301	1.440	1.667	0.9993
ARCH Lag[7]	0.02004	2.315	1.543	1.0000

Table 4.58 represents the lags of ARCH model with stastistic of ARCH lag (3)=0.0599, ARCH lag (5)= 0.01301 and ARCH lag (7)=0.02004. The weighted ARCH LM test showed that the p-values for the ARCH lags are insignificant, p-vale is greater than 5% level of significant.

Table 4.59: ARCH model of water evaporation for Ga-Rantho dam.

Weighted ARCH LM Tests				
	Statistic	Shape	Scale	P-Value
ARCH Lag[3]	0.00599	0.500	2.000	0.9383
ARCH Lag[5]	0.01301	1.440	1.667	0.9993
ARCH Lag[7]	0.02004	2.315	1.543	1.0000

Table 4.59 represents the lags of ARCH model with stastistic of ARCH lag (3)=0.0599, ARCH lag (5)= 0.01301 and ARCH lag (7)=0.02004. The weighted ARCH LM test showed that the p-values for the ARCH lags are insignificant, p-vale is greater than 5% level of significant.

Table 4.60: ARCH model of water evaporation for Leeukraal DeHoop dam.

Weighted ARCH LM Tests				
	Statistic	Shape	Scale	P-Value
ARCH Lag[3]	0.02441	0.500	2.000	0.8758
ARCH Lag[5]	0.05861	1.440	1.667	0.9938
ARCH Lag[7]	0.08724	2.315	1.543	0.9995

Table 4.60 represents the lags of ARCH model with stastistic of ARCH lag (3)=0.0244, ARCH lag (5)= 0.05861 and ARCH lag (7)=0.08724. The weighted ARCH LM test showed that the p-values for the ARCH lags are insignificant, p-vale is greater than 5% level of significant.

Table 4.61: ARCH model of water evaporation for Luphephe dam.

Weighted ARCH LM Tests				
	Statistic	Shape	Scale	P-Value
ARCH Lag[3]	0.00599	0.500	2.000	0.9383
ARCH Lag[5]	0.01301	1.440	1.667	0.9993
ARCH Lag[7]	0.02004	2.315	1.543	1.0000

Table 4.61 represents the lags of ARCH model with stastistic of ARCH lag (3)=0.0599, ARCH lag (5)= 0.01301 and ARCH lag (7)=0.02004. The weighted ARCH LM test showed that the p-values for the ARCH lags are insignificant, p-vale is greater than 5% level of significant.

Table 4.62: ARCH model standardised residuals of water evaporation for Mokolo dam.

Standardised Residuals Tests				
Statistic				p-value
Ljung-Box Test	R	Q(10)	1.471601	0.9990211
Ljung-Box Test	R	Q(15)	2.306023	0.9999245
Ljung-Box Test	R	Q(20)	2.984228	0.9999961
Ljung-Box Test	R^2	Q(10)	0.01039123	1
Ljung-Box Test	R^2	Q(15)	0.01497463	1
Ljung-Box Test	R^2	Q(20)	0.01940608	1
LM Arch Test	R	TR^2	0.01249917	1

Table 4.62 present standardised residual for ARCH model using Ljung-Box test with results of an insignificant residual p-value and a squared residual p-value, the p-values are higher than 5% level of significant and the LM-ARCH test failed to reject the null hypothesis. The residual values are (1.472, 2.306, 2.984) and squared residual value are (0.0104, 1.0149,0.01941, 0.0125) (Bollerslev, 1987; Fathian, 2019).

Table 4.63: ARCH model standardised of water evaporation for Ga-Rantho dam.

Standardised Residuals Tests			
Statistic			P-value
Ljung-Box Test	R	Q(10)	0.6367242
Ljung-Box Test	R	Q(15)	1.4867
Ljung-Box Test	R	Q(20)	1.940654
Ljung-Box Test	R^2	Q(10)	0.01145807
Ljung-Box Test	R^2	Q(15)	0.01654399
Ljung-Box Test	R^2	Q(20)	0.02166026
LM Arch Test	R	TR^2	0.01380413

Table 4.63 present standardised residual for ARCH model using Ljung-Box test with results of an insignificant residual p-value and a squared residual p-value, the p-values are higher than 5% level of significant and the LM- ARCH test failed to reject the null hypothesis. The residual values are (0.637, 1.487, 1.941) and squared residual value are (0.0115, 0.0217, 0.0138) (Bollerslev, 1987; Fathian, 2019).

Table 4.64: ARCH model standardised residuals of water evaporation for Leeukraal DeHoop dam.

Standardised Residuals Tests				
Statistic				p-value
Ljung-Box Test	R	Q(10)	1.493252	0.9989562
Ljung-Box Test	R	Q(15)	2.190411	0.999946
Ljung-Box Test	R	Q(20)	2.360534	0.9999995
Ljung-Box Test	R^2	Q(10)	0.007298613	1
Ljung-Box Test	R^2	Q(15)	0.01051426	1
Ljung-Box Test	R^2	Q(20)	0.01399903	1
LM Arch Test	R	TR^2	0.008768596	1

Table 4.64 present standardised residual for ARCH model using Ljung-Box test with results of an insignificant residual p-value and a squared residual p-value, the p-values are higher than 5% level of significant and the LM- ARCH test failed to reject the null hypothesis. The residual values are (1.493, 2.190, 2.361) and squared residual values are (0.0073, 0.0105, 0.0139) (Bollerslev, 1987; Fathian, 2019).

Table 4.65: ARCH model standardised residuals of water evaporation for Luphephe dam.

Standardised Residuals Tests				
Statistic				p-value
Ljung-Box Test	R	Q(10)	2.1627	$1.409201e-09$
Ljung-Box Test	R	Q(15)	2.332259	$2.999659e-08$
Ljung-Box Test	R	Q(20)	3.17977	$3.600325e-07$
Ljung-Box Test	R^2	Q(10)	0.051627	0.9906365
Ljung-Box Test	R^2	Q(15)	0.031274	0.9997197
Ljung-Box Test	R^2Q	(20)	0.0490861	0.9947361
LM Arch Test	R	TR^2	2.77636	0.9969275

Table 4.65 present standardised residual for ARCH model using Ljung-Box test with results of an insignificant residual p-value and a squared residual p-value, the p-values are higher than 5% level of significant and the LM- ARCH test failed to reject the null hypothesis. The residual values are (2.163, 2.332, 3.179) and squared residual values are (0.0516, 0.0313, 0.0491) (Bollerslev, 1987; Fathian, 2019).

4.2.8 GARCH model of water evaporation from the selected dams

Table 4.66: GARCH model of water evaporation for Mokolo dam.

Parameters				
	Estimate	Std. Error	t value	Pr(> |t|)
mu	4.940159	0.668295	7.3922	0
AR1	0.993163	0.002409	412.2411	0
MA1	-0.459469	0.024506	-18.7495	0
omega	0.057577	0.010845	5.3093	0
alpha1	0.439953	0.033397	13.1733	0
beta1	0.559047	0.033248	16.8146	0
shape	3.179263	0.101054	31.4610	0

Table 4.66 present parameter estimations for GARCH model, the mean estimate value is 4.940159 , autoregressive (1) estimate value of 0.993163 and a moving avrage (1) estimate value of -0.459469. The p-value of the mean, autoregressive and moving average estimates are significant, the p-values are less than 5% level of significant (Matringe and Guida, 2004; Stehlikova, 2005).

Table 4.67: GARCH model parameter estimation for Ga-Rantho dam.

Parameters				
	Estimate	Std. Error	t value	Pr(> |t|)
mu	4.940159	0.668295	7.3922	0
AR1	0.993163	0.002409	412.2411	0
MA1	-0.459469	0.024506	-18.7495	0
omega	0.057577	0.010845	5.3093	0
alpha1	0.439953	0.033397	13.1733	0
beta1	0.559047	0.033248	16.8146	0
shape	3.179263	0.101054	31.4610	0

Table 4.67 presents parameter estimations for GARCH model, the mean estimate value is 4.940159 , autoregressive (1) estimate value of 0.993163 and a moving avrage (1) estimate value of -0.459469. The p-value of the mean, autoregressive and moving average estimates are significant, the p-values are less than 5% level of significant (Matringe and Guida, 2004; Stehlikova, 2005).

Table 4.68: GARCH model parameters of water evaporation for Leeukraal DeHoop dam.

Parameters				
	Estimate	Std. Error	t value	Pr($>\|t\|$)
mu	4.940159	0.408996	12.0787	0.000000
AR1	0.993163	0.002921	339.9869	0.000000
MA1	-0.459469	0.041834	-10.9832	0.000000
omega	0.057577	0.041235	1.3963	0.162615
alpha1	0.439953	0.078834	5.5807	0.000000
beta1	0.559047	0.131352	4.2561	0.000021
shape	3.179263	0.102410	31.0443	0.000000

Table 4.68 presents parameter estimations for GARCH model, the mean estimate value is 4.940159 , autoregressive (1) estimate value of 0.993163 and a moving avrage (1) estimate value of -0.459469. The p-value of the mean, autoregressive and moving average estimates are significant, the p-values are less than 5% level of significant (Matringe and Guida, 2004; Stehlikova, 2005).

Table 4.69: GARCH model parameter of water evaporation for Luphephe dam.

Parameters				
	Estimate	Std. Error	t value	Pr($>\|t\|$)
mu	$2.0151e^{+03}$	0.124368	$1.6203e^{+04}$	0.00000
AR1	$1.0000e^{+00}$	0.000038	$2.6368e^{+04}$	0.00000
MA1	$3.6613e^{-01}$	0.009029	$4.0551e^{+01}$	0.00000
omega	$0.0000e^{+00}$	0.000000	$1.4872e^{-02}$	0.98813
alpha1	$3.6635e^{-01}$	0.023271	$1.5743e^{+01}$	0.00000
beta1	$9.7965e^{-02}$	0.013215	$7.4129e^{+00}$	0.00000
shape	$2.1156e^{+00}$	0.003459	$6.1159e^{+02}$	0.00000

Table 4.69 presents parameter estimations for GARCH model, the mean estimate value is 2.0151 $e + 03$, autoregressive (1) estimate value of 1.00000 $e + 00$and a moving avrage (1) estimate value of $3.6613e - 01$. The p-value of the mean, autoregressive and moving average estimates are significant, the p-values are less than 5% level of significant (Matringe and Guida, 2004; Stehlikova, 2005).

4.2.9 GARCH (1,1) volatility plot for the selected dams

Figure 4.25 presents the volatility plot for GARCH model, the plot shows a low constant water evaporation variation pattern in the plot for several years, then years 2016 and 2017 the water evaporation rate plot pattern had a high water evaporatuion variation (Stehlikova, 2005).

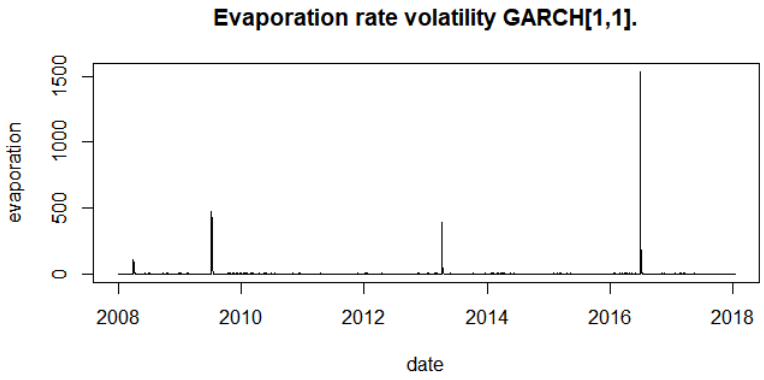


Figure 4.25: Volatility plot Mokolo dam

Figure 4.26 presents the volatility plot for GARCH model, the pattern of the plot has low variation of water evaporation rate for several years. The water evaporation rate highly varies in the year 2014 (Stehlikova, 2005).

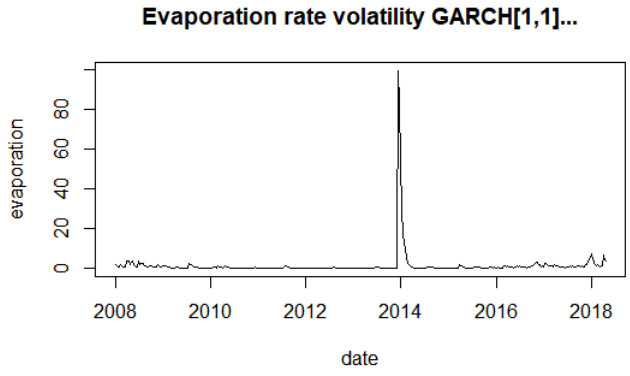


Figure 4.26: Volatility plot Ga-Rantho dam.

Figure 4.27 presents the volatility plot for GARCH model, the pattern of the plot have low variation of water evaporation rate in the other years of the plot, where as the water evaporation rate highly varies in the year 2017 and 2018 (Stehlikova, 2005).

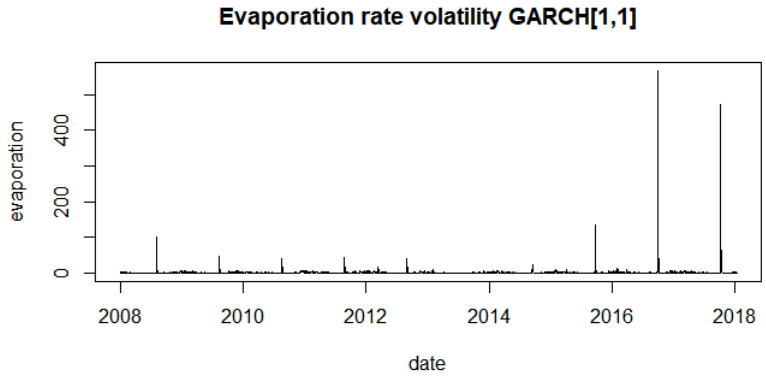


Figure 4.27: Volatility plot Leeukraal DeHoop dam.

Figure 4.28 presents the volatility plot for GARCH model, the pattern of the plot have low variation of water evaporation rate in other year of the plot, where water evaporation rate highly varies in the year 2014 (Stehlikova, 2005).

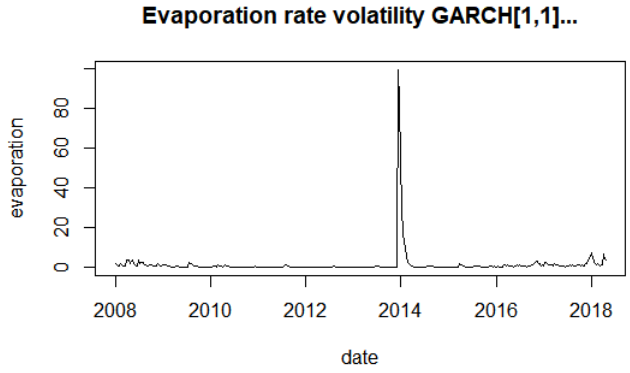


Figure 4.28: Volatility plot Luphephe dam.

4.2.10 Dignostic checking

Figure 4.29 presents that there are no significant peaks on the residual ACF and also on the residual PACF plot. The plots show that the residuals are white noise since the peaks are insignificant.

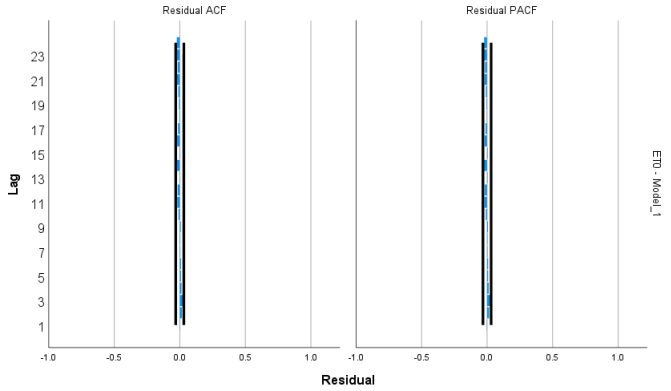


Figure 4.29: Residual plot for ACF and PACF for ARIMA (1,1,1) model for water evaporation Mokolo dam.

Figure 4.30 presents the Q-Q plot for ARIMA (1,1,1) model, most of the points on the plot fall on the line and there is a linear pattern, hence there is a normal distribution. There are a couple of outliers on the far right and a couple of outliers on the far left (Stine, 2017).

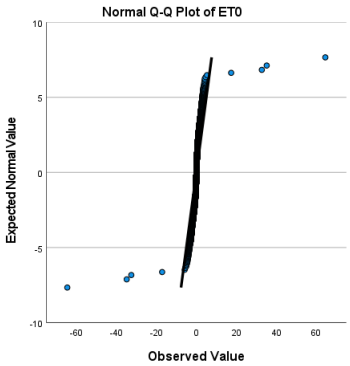


Figure 4.30: Q-Q plot for ARIMA(1,1,1) model for water evaporation Mokolo damv

Figure 4.31 presents that there are no significant peaks on the residual ACF and also on the residual PACF plot. The plots show that the residuals are white noise since the peaks are insignificant.

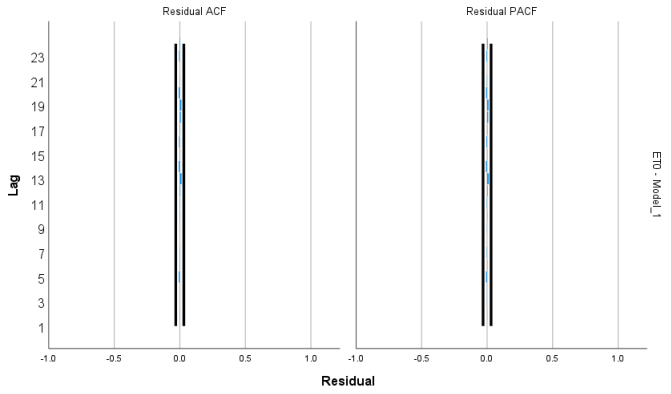


Figure 4.31: Residual plot for ACF and PACF for ARIMA (1,1,2) model for water evaporation Ga-Rantho dam.

Figure 4.32 presents the Q-Q plot for ARIMA (1,1,2) model, most of the points on the plot fall on the line and there is a linear pattern, hence there is a normal distribution. There are a couple of outliers on the far right and a couple of outliers on the far left (Stine, 2017).

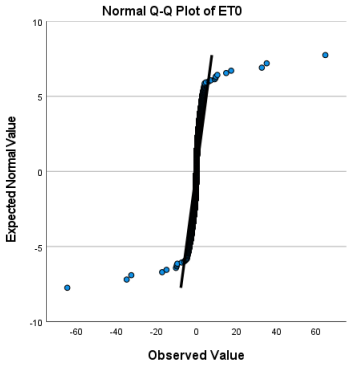


Figure 4.32: Q-Q plot for ARIMA (1,1,2) model for water evaporation Ga-Rantho dam.

Figure 4.33 presents that there are no significant peaks on the residual ACF and also on the residual PACF plot. The plots show that the residuals are white noise since the peaks are insignificant.

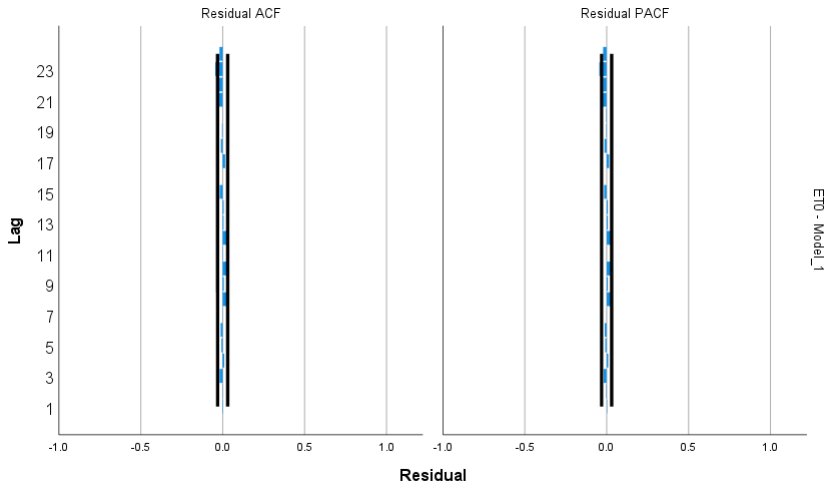


Figure 4.33: Residual plot for ACF and PACF for ARIMA (1,1,1) model for water evaporation Leeukraal DeHoop dam.

Figure 4.34 presents the Q-Q plot for ARIMA (1,1,3) model, most of the points on the plot fall on the line and there is a linear pattern, hence there is a normal distribution. There are a couple of outliers on the far right and a couple of outliers on the far left (Stine, 2017).

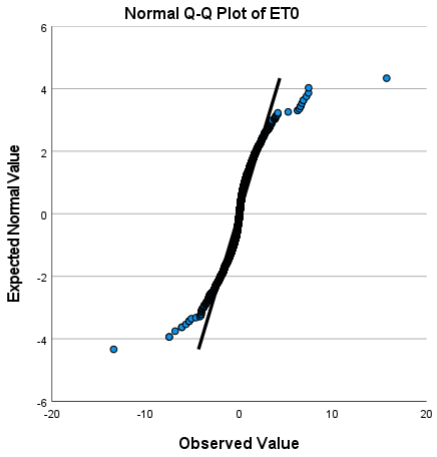


Figure 4.34: Q-Q plot for ARIMA (1,1,3) model for water evaporation Leeukraal DeHoop dam.

Figure 4.35 presents that there are no significant peaks on the residual ACF and also on the residual PACF plot. The plots show that the residuals are white noise since the peaks are insignificant.

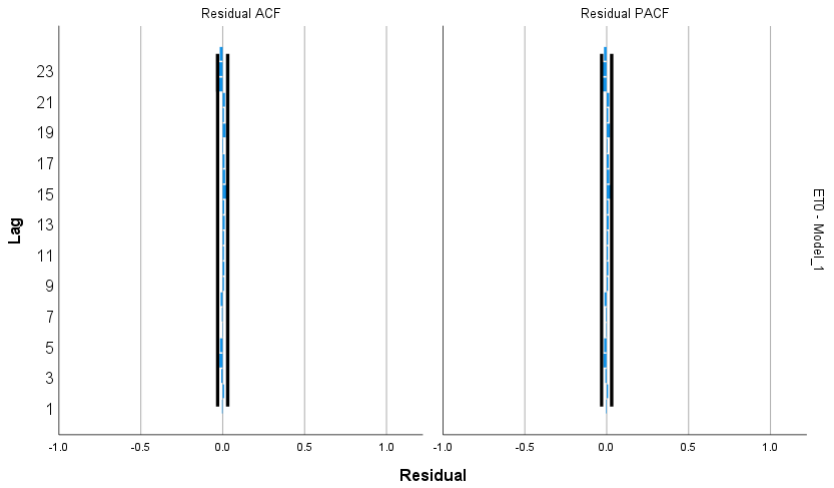


Figure 4.35: Residual plot for ACF and PACF for ARIMA (1,1,3) model for water evaporation Luphephe dam.

Figure 4.36 presents the Q-Q plot for ARIMA (1,1,3) model, most of the points on the plot fall on the line and there is a linear pattern, hence there is a normal distribution. There are a couple of outliers on the far right and a couple of outliers on the far left (Stine, 2017).

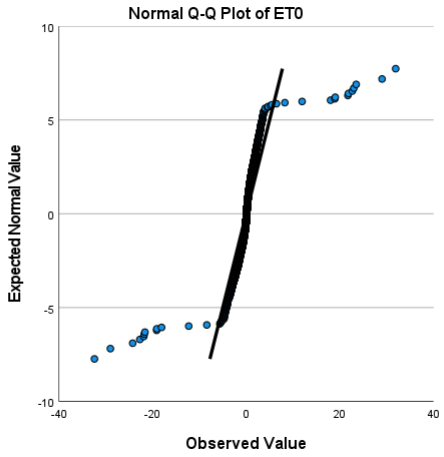


Figure 4.36: Q-Q plot for ARIMA (1,1,3) model for water evaporation Luphephe dam.

Figure 4.37 presents that there are no significant peaks on the residual ACF and also on the residual PACF plot. The plots show that the residuals are white noise since the peaks are insignificant.

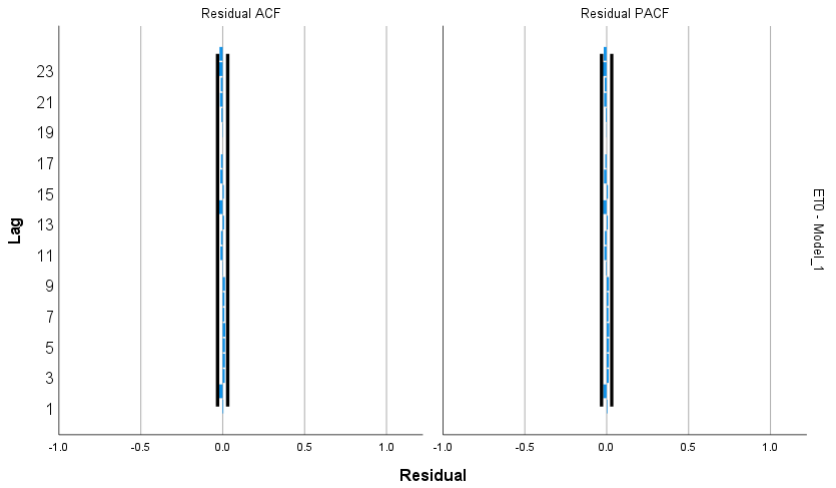


Figure 4.37: Residual plot for ACF and PACF for ARIMA (1,1,2) model for water evaporation Mokolo dam.

Figure 4.38 presents the Q-Q plot for ARIMA (1,1,2) model, most of the points on the plot fall on the line and there is a linear pattern, hence there is a normal distribution. There are two outliers on the far left and a couple of outliers on the very far right (Stine, 2017).

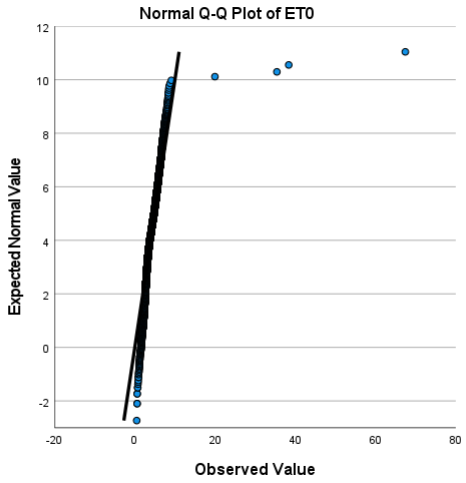


Figure 4.38: Q-Q plot for ARIMA (1,1,2) model for water evaporation Mokolo dam.

Figure 4.39 presents that there are no significant peaks on the residual ACF and also on the residual PACF plot. The plots show that the residuals are white noise since the peaks are insignificant.

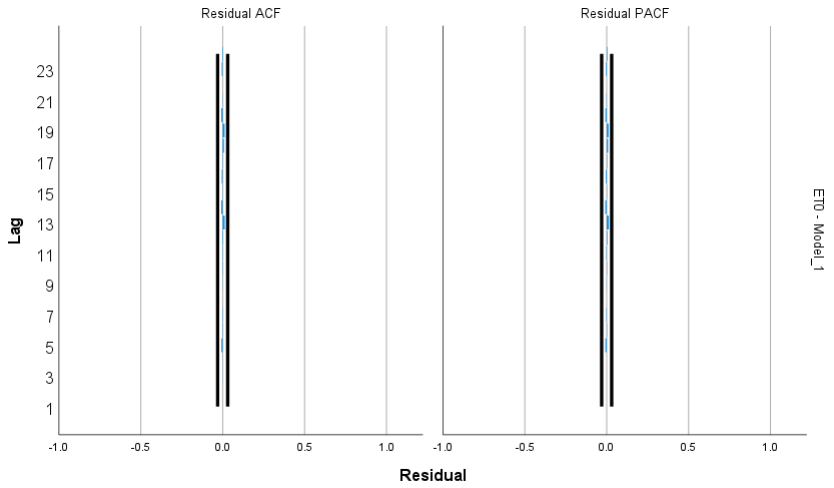


Figure 4.39: Residual plot for ACF and PACF for ARIMA (1,1,3) model for water evaporation Ga-Rantho dam.

Figure 4.40 presents the Q-Q plot for ARIMA (1,1,3) model, most of the points on the plot fall on the line and there is a linear pattern, hence there is a normal distribution. There are two outliers on the far left and a couple of outliers on the very far right (Stine, 2017).

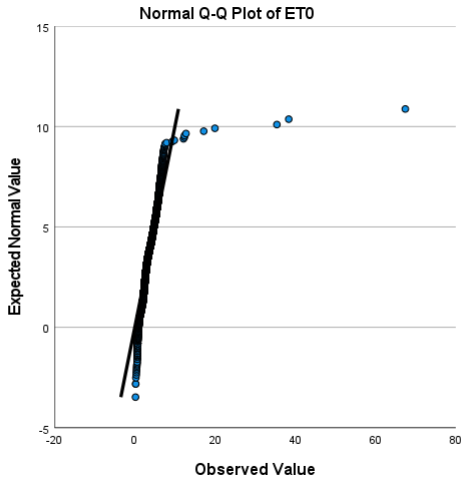


Figure 4.40: Q-Q plot for ARIMA (1,1,3) model for water evaporation Ga-Rantho dam.

Figure 4.41 presents that most of the peaks are insignificant from lag 1 to lag 21 and on lag 23 the peaks is significant on the residual ACF and also on the residual PACF plot . The plots show that the residuals are white noise since most of the peaks are insignificant.

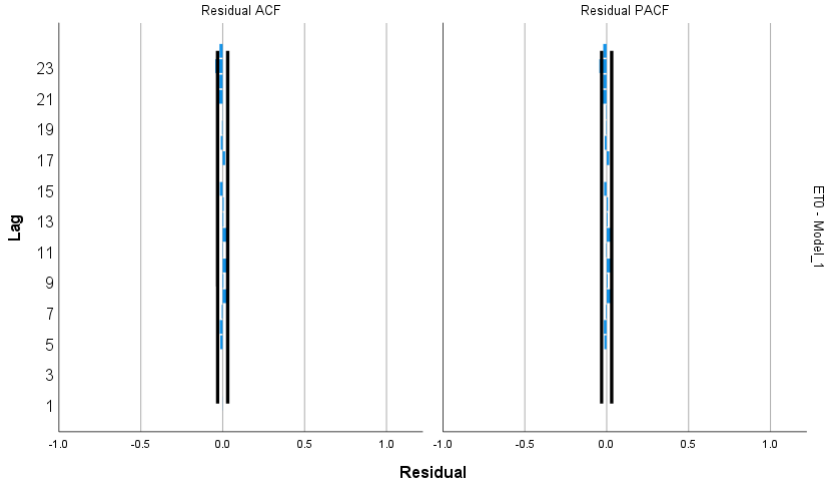


Figure 4.41: Residual plot for ACF and PACF for ARIMA (1,1,3) model for water evaporation Leeukraal DeHoop dam.

Figure 4.42 presents the Q-Q plot for ARIMA (1,1,3) model, some of the points on the plot fall on the line and there is a curve pattern, hence it is not a normal distribution. There are a couple of outliers on the far right and a couple of outliers on the far left (Stine, 2017).

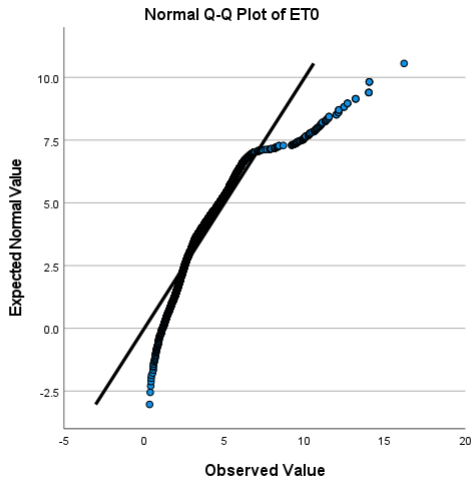


Figure 4.42: Q-Q plot for ARIMA (1,1,3) model for water evaporation Leeukraal DeHoop dam.

Figure 4.43 presents that there are no significant peaks on the residual ACF and also on the residual PACF plot. The plots show that the residuals are white noise since the peaks are insignificant.

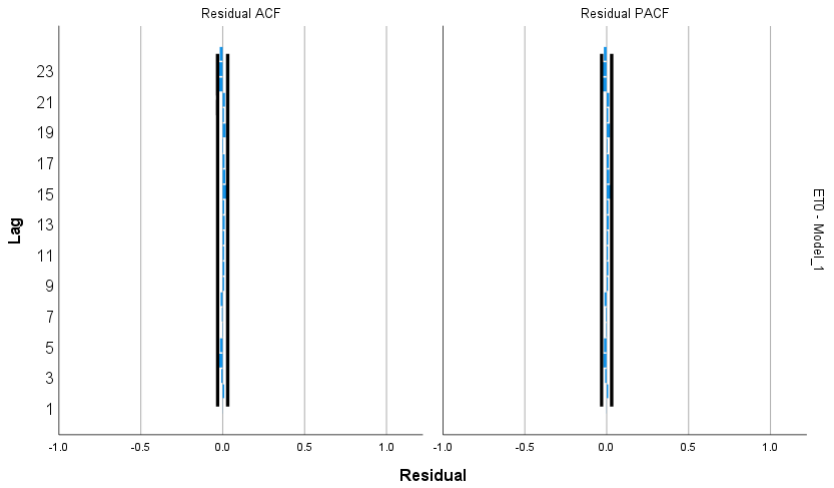


Figure 4.43: Residual plot for ACF and PACF for ARIMA (2,1,2) model for water evaporation Luphephe dam.

Figure 4.44 presents the Q-Q plot for ARIMA (2,1,2) model, most of the points on the plot fall on the line and there is a linear pattern, hence there is a normal distribution. There are a couple of outliers on the far right and a couple of outliers on the far left (Stine, 2017).

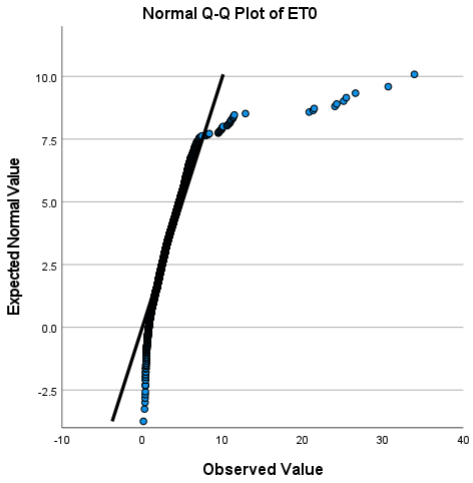


Figure 4.44: Q-Q plot for ARIMA (2,1,2) model for water evaporation Luphephe dam.

4.2.11 Model varification of ARIMA (p,d,q) model of the selected dams

Table 4.70, table 4.71, table 4.72 and table 4.73 presents selected models with results from the selected dams. The best model will be selected from each selected dams, which will be used to forecast and predict water evaporation from each selected dams.

Table 4.70: Model verification of ARIMA (p,d,q) models Mokolo dam.

Model	BIC	RMSE	MAPE	MAE	Fit statistics
ARIMA (1,1,1)	1.039	1.674	22.941	0.778	13.315
ARIMA (1,1,2)	0.872	0.524	19.023	0.730	11.654
ARIMA (2,1,3)	1.042	1.672	23.019	0.782	5.094

Table 4.71: Model verification of ARIMA (p,d,q) models Ga-Rantho dam.

Model	BIC	RMSE	MAPE	MAE	Fit statistics
ARIMA (1,1,2)	0.139	0.543	20.283	1.733	1.930
ARIMA (1,1,3)	1.041	1.673	25.280	0.733	1.896
ARIMA (2,1,2)	1.041	1.673	25.282	0.733	1.898

Table 4.72: Model verification of ARIMA (p,d,q) models Leeukraal DeHoop dam.

Model	BIC	RMSE	MAPE	MAE	Fit statistics
ARIMA (1,1,1)	0.171	0.385	18.027	0.706	17.394
ARIMA (1,1,3)	0.175	1.085	24.014	0.706	15.273
ARIMA (2,1,1)	0.173	1.085	24.039	0.506	17.216

Table 4.73: Model verification of ARIMA (p,d,q) models Luphephe dam.

Model	BIC	RMSE	MAPE	MAE	Fit statistics
ARIMA (1,1,3)	1.107	1.728	55.961	1.016	10.792
ARIMA (2,1,2)	1.109	1.731	55.843	1.010	5.122
ARIMA (2,1,3)	0.113	0.479	50.966	0.416	11.203

4.2.12 ARIMA forecasting of water evaporation rate for the selected dam

Figure 4.45 presents the forecasting of water evaporation rate in Mokolo dam. The black pattern of the plot represent the original time series plot from year 2008 to 2018, the blue line of the plot represents the forecasting of water evaporation for a period of 3 years. The results show a constant pattern of water evaporation rate (Bari et al., 2015).

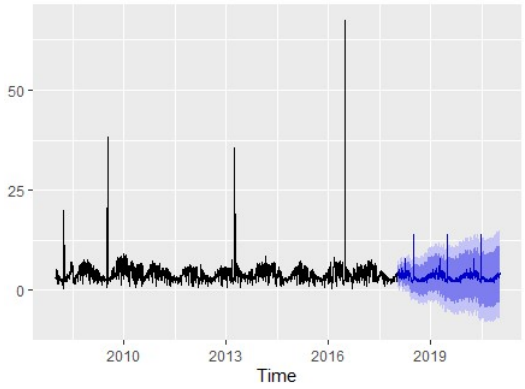


Figure 4.45: Water evaporation rate Mokolo dam.

Figure 4.46 presents the forecasting of water evaporation rate in Ga-Rantho dam. The black pattern of the plot represent the original time series plot from year 2008 to 2018, the blue line of the plot represents the forecasting of water evaporation for a period of 3 years. The results show a constant increase pattern of water evaporation rate (Bari et al., 2015).

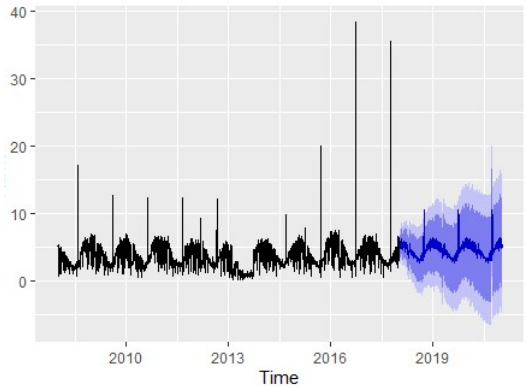


Figure 4.46: Water evaporation rate Ga-Rantho dam.

Figure 4.47 presents the forecasting of water evaporation rate in Leeukraal DeHoop dam. The black pattern of the plot represent the original time series plot from year 2008 to 2018, the blue line of the plot represents the forecasting of water evaporation for a period of 3 years. The results show a constant increase pattern of water evaporation rate (Bari et al., 2015).

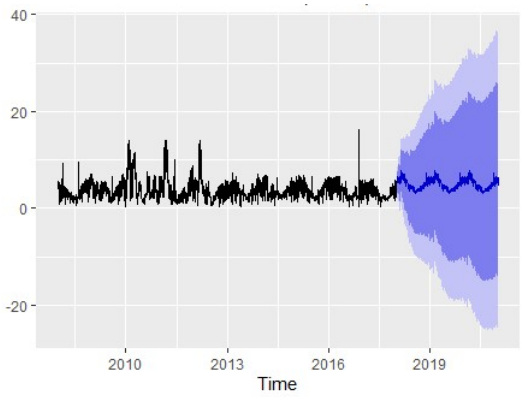


Figure 4.47: Water evaporation rate Leeukraal DeHoop dam.

Figure 4.48 presents the forecasting of water evaporation rate in Luphephe dam. The black pattern of the plot represent the original time series plot from year 2008 to 2018, the blue line of the plot represents the forecasting of water evaporation for a period of 3 years. The results show a constant increase pattern of water evaporation rate (Bari et al., 2015).

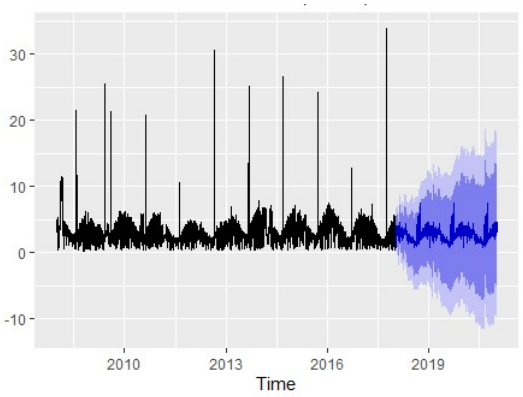


Figure 4.48: Water evaporation rate Luphephe dam.

Figure 4.49 presents a forecasting time series plot of water evaporation rate in Mokolo dam. The plot shows the original time series plot in black pattern line, where the blue line indicates the forecast of the best selected ARIMA model for the period of 3 years, from year 2019 to 2021. The best selected model used to forecast the water evaporation rate is ARIMA (1,1,2). The forecasted water evaporation rate for the period of five years has indicated a constant increase of water evaporation rate (Bari et al., 2015).

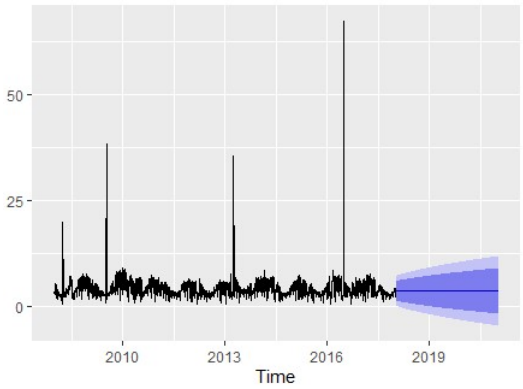


Figure 4.49: Forcating time series plot of water evaporation rate Mokolo dam.

Figure 4.50 presents a forecasting time series plot of water evaporation rate in Ga-rantho dam. The plot shows the original time series plot in black pattern line, where the blue line indicates the forecast of the best selected ARIMA model for the period of 3 years, from year 2019 to 2021. The best selected model used to forecast the water evaporation rate is ARIMA (1,1,2). The forecast water evaporation rate for the period of five years has indicated a constant rate of water evaporation rate (Bari et al., 2015).

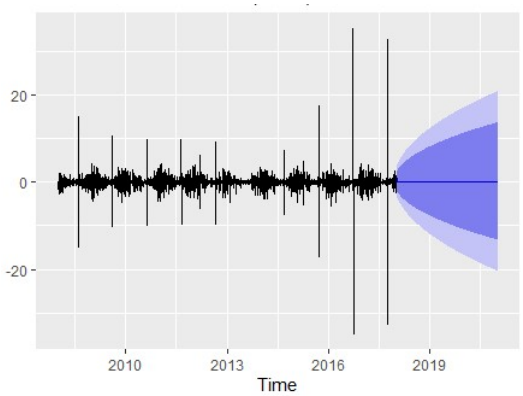


Figure 4.50: Forcating time series plot of water evaporation rate Ga-Rantho dam.

Figure 4.51 presents a forecasting time series plot of water evaporation rate in Leeukraal DeHoop dam. The plot shows the original time series plot in black pattern line, where the blue pattern line indicates the forecast of the best selected ARIMA model for the period of 3 years, from year 2019 to 2021. The best selected model used to forecast the water evaporation rate is ARIMA (1,1,1). The forecast water evaporation rate for the period of five years has indicated a constant increase of water evaporation rate with zero mean rate(Bari et al., 2015).

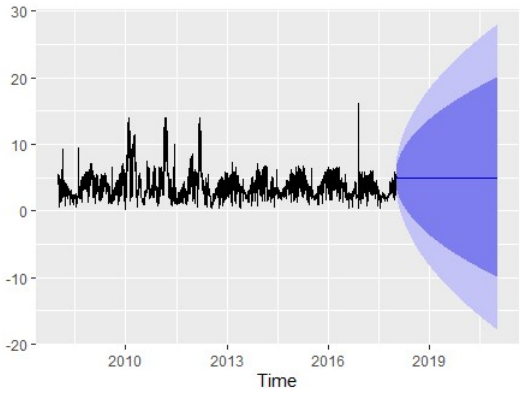


Figure 4.51: Forcating time series plot of water evaporation rate Leeukraal DeHoop dam.

Figure 4.52 presents a forecasting time series plot of water evaporation rate in Luphephe dam. The plot shows the original time series plot in black pattern line, where the blue pattern line indicates the forecast of the best selected ARIMA model for the period of 3 years, from year 2019 to 2021. The best selected model used to forecast the water evaporation rate is ARIMA (2,1,3). The forecast water evaporation rate for the period of five years has indicated a constant increase of water evaporation rate (Bari et al., 2015).

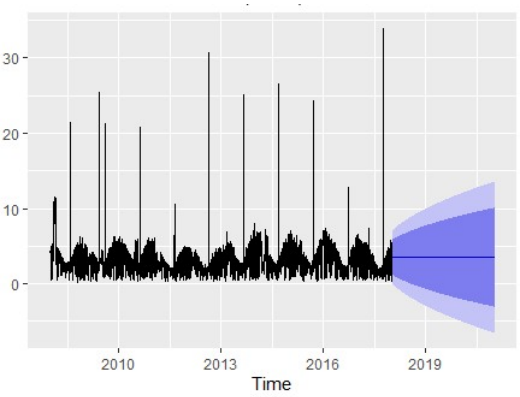


Figure 4.52: Forecasting time series plot of water evaporation rate Luphephe dam.

Chapter 5

Conclusion and Recommendations

5.1 Introduction

This chapter presents the conclusion and recommendations on the statistical analysis of water evaporation from the selected dams in the Limpopo province South Africa. The conclusion is presented in the first part of the chapter with concluding outcomes based on chapter 4. The second part presents recommendations based on the results in the study.

5.2 Conclusion

The study sets out to perform time series modelling of evaporation from selected dams in the Limpopo province of South Africa. The five objectives mentioned in chapter 1 were mentioned, which were looked into. Using the daily water evaporation rate time series data from the selected dams in the Limpopo province of South Africa from year 2008 to year 2018, the time series models such as ARIMA model to forecast, ARCH, GARCH and VAR models were investigated. In the study a time series plot was done for checking if there was a trend variation, seasonal variation, irregular or cyclical variation. The time series plots revealed that the time series data has a trend , hence a test was carried out to investigate if the time series data used was a stationary or non-stationary time series data. The augmented Dickey-Fuller

test was used, and it revealed that the time series is non-stationary for all the selected dams at 5% level of significance. Since the time series data was non-stationary a method called differencing was used to convert the data into a stationary time series data, hence after the method was applied the time series data was then stationary. The augmented Dicky-Fuller test was used to check the stationarity of the differenced time series data. The ADF results showed that the time series data was stationary at 5% level of significant for all the dams in the study.

The autocorrelation function and partial autocorrelation function plots were used to construct an ARIMA(p, d, q) model, whereby the plots revealed that few lags where significant to construct an ARIMA (p, d, q) model of water evaporation for the selected dams. The model fit and model statistics revealed that ARIMA model of order (1,1,2) for Mokolo dam with a RMSE of 0.524 ,ARIMA mode of order (1, 1, 2) for Ga-Rantho with RMSE of 0.543 , Leeukraal DeHoop dam ARIMA model of order (1, 1, 1) with RMSE of 0.385 and Luphephe dam of ARIMA of order (2, 1, 3) with RMSE of 0.479 showed that the model can relatively predict water evaporation time series data accuracy. The BIC value of the ARIMA models from the selected dams reveals that the ARIMA (1, 1, 2), ARIMA (1, 1, 2), ARIMA (1, 1, 1) and ARIMA (2, 1 3) are the best models of all the models selected from the dams, since the value of BIC was small.

The Ljung-Box test reveals that the statistic for Mokolo dam ARIMA(1, 1, 2) , Leeukraal DeHoop dam ARIMA (1, 1, 1) are insignificant at 5% level of significant and Ga-Rantho dam ARIMA (1, 1, 2), Luphephe dam ARIMA (2, 1, 3) are significant at 5% level of significant, this means there is an increase or decrease of water evaporation rate from the selected dams.The parameter estimation reveals that water evaporation from the selected dams in Limpopo province of South Africa is to increase or decrease due to the contributing factors such as rain, temperature, and humidity. The estimated parameters of the selected damns showed significance and insignificance of the models at 5% level of significant.

The VAR model was used to determine the relationship between evaporation, temperature, and Rainfall. The VAR(10) model provides an adequate representation of evaporation, rainfall, and temperature, the coefficients estimates and p-value of less than 5% level of significant shows that the model is significant , hence there is an existing relationship between Evaporation and the explanatory variables. The VAR (10) correlation residuals revealed positive relationship between evaporation and temperature. There is a weak

positive relationship between evaporation and rain, hence there is a high chance that evaporation rate is to increase.

The standardised residuals of the selected dams (Mokolo dam, Leeukraal DeHoop dam and Luphephe dam) reveals that the p-values are higher than 5% level of significant of the LjungBox test residual and square residuals , therefore the null hypothesis cannot be rejected neither in residual nor in squared residuals. The insignificant p-values does not gives a good way to obtain autocorrelation and time-vary volatility in the water evaporation time series data. The LM-ARCH test failed to reject the null hypothesis of absence of ARCH effects. The standardised residual for Ga-Rantho dam reveals insignificant residual p-values greater than 5% level of significant and a significant squared residual, the null hypothesis was rejected for the squared residual.

The GARCH (1,1) model reveals that there was no heteroscedasticity , Ljung Box test presented weak evidence of ARCH effect in the residuals water evaporation time series, this recommend that the GARCH (1,1) modelling volatility cannot be used.The parameter revealed that GARCH model is significant with a p-value less than 5% level of significant. The volatility plot for the selected dams showed that the GARCH (1,1) model have low ability to model water evaporation rate values, under the influence of the contributing factors such as temperature, rainfall, and humidity. The GARCH (1,1) model is not able to influence and predict water evaporation rate. The volatility results revealed that in the year 2016 and 2017 there was high variation of water evaporation rate in Mokolo dam, Ga-Rantho dam showed that water evaporation rate had a high variation in 2014, Leeukraal DeHoop dam showed that water evaporation rate has a high variation in 2017 and 2018 and Luphephe dam showed that water evaporation rate had a high variation in 2014. Where water evaporation rate varies indicates that tempeature is very hight and there is minimum rainfall.

The ACF and PACF revealed that the selected ARIMA models have no significant peaks on the ACF and PACF residuals the models are insignificant from the selected dam, the Q-Q plot showed that the water evaporation time series data does not have outliers and is normally distributed. The best models from each dam were selected by taking the smallest value of BIC. The best ARIMA models selected from each dam were used to forecast the water evaporation rate from the selected dam. The forecast revealed that water evaporation rate Mokolo dam after 3 years will be constant, Ga-Rantho dam after 3 years showed that there will be a hyperbolically decrease

in water evaporation rate, Leeukraal DeHoop dam showed that after 3 years there will be a constant value of water evaporation rate and Luphephe dam showed a bit of increase then a constant value of water evaporation rate in the next 3 years. The forecasting results shows that there will not be much of an impact to the dams and the communities, where water can be impacted by the rate at which evaporation might occur. Modelling and forecasting of water evaporation rate in the selected dams will improve the process taken to control the rate of water evaporation.

5.3 Recommendations

Based on the findings of the study of time series modelling of water evaporation from the selected dams in the Limpopo province of South Africa. More dams can be constructed for the future to prevent the chance of experiencing shortage of water. The water evaporation rate might not look like a threat now, but as the temperature increases and having less rainfall this can cause a major problem in the Limpopo province communities and South Africa as a whole. The South African Weather Service should take into consideration the regular data capturing of water evaporation rate and implementing future dams constructions for water security.

This research suggests future research that can help into managing the possibility of a future increase in water evaporation rate. The study revealed that there will be a constant positive rate of evaporation rate from the selected dams in the Limpopo province of South Africa in the future based on the forecasted models. The study recommends that factors contributing towards the occurrence of water evaporation to be modelled using time series models, more especial SARIMA model.

References

Adhikari, R. and Agrawal, R.K., 2013. An introductory study on time series modelling and forecasting. *arXiv preprint arXiv*:1302.6613.

Ali Ghorbani, M., Kazempour, R., Chau, K.W., Shamshirband, S. and Taherei Ghazvinei, P., 2018. Forecasting pan evaporation with an integrated artificial neural network quantum-behaved particle swarm optimization model: a case study in Talesh, Northern Iran. *Engineering Applications of Computational Fluid Mechanics*, **12**(1), 724-737.

Allawi, M.F., Binti Othman, F., Afan, H.A., Ahmed, A.N., Hossain, M., Fai, C.M. and El-Shafie, A., 2019. Reservoir evaporation prediction modelling based on artificial intelligence methods. *Water*, **11**(6), 1226.

Althoff, D., Rodrigues, L.N. and da Silva, D.D., 2019. Evaluating evaporation methods for estimating small reservoir water surface evaporation in the Brazilian savannah. *Water*, **11**(9), 1942.

Arbain, S.H. and Wibowo, A., 2012.Neural networks based nonlinear time series regression for water level forecasting of Dungun River.*Journal of Computer Science*, **8**(9), 1506.

Babazadeh, H. and Shamsnia, S.A., 2014. Modelling climate variables using time series analysis in arid and semi arid regions. *African Journal of Agricultural Research*, **9**(26), 2018-2027.

Bakar, N.A. and Rosbi, S., 2017. Data Clustering using Autoregressive Integrated Moving Average (ARIMA) model for Islamic Country Currency: An Econometrics method for Islamic Financial Engineering. *The International Journal of Engineering and Science (IJES)*, **6**(6), 22-31.

Bari, S.H., Rahman, M.T., Hussain, M.M. and Ray, S., 2015. Forecasting monthly precipitation in Sylhet city using ARIMA model. *Civil and Environmental Research*, **7**(1), pp.69-77.

Baydaronglu, O. and Kocak, K., 2014. Reconstruction of Evaporation Dynamics from Time Series. *Chaotic Modelling and Simulation (CMSIM)*, 3: 271-279

Behrouzil K. and Chini S.F., 2017. Evaluation of Evaporation Estimation Methods: A Case Study of Karaj Dam Lake. *Journal of Computational Applied Mechanics*, **48**(1), 137-150.

Bernal, J.L., Soumerai, S. and Gasparrini, A., 2018. A methodological framework for model selection in interrupted time series studies. *Journal of clinical epidemiology*, **103**, 82-91.

Bollerslev, T., 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, 542-547.

Botai C.M., Botai J.O. and Adeola A.M., 2018. Spatial distribution of temporal precipitation contrasts in South Africa. *South African Journal of Science*, **114**(7-8), 70-78.

Brownlee, J., 2017. *Introduction to time series forecasting with python: how to prepare data and develop models to predict the future*. Machine Learning Mastery.

Brutsaert W., 2013. *Evaporation into the atmosphere: theory, history and applications* **(Vol.1)**. Springer Science & Business Media(Germany).

Burns, P., 2002. Robustness of the Ljung-Box test and its rank equivalent. *Available at SSRN 443560*.

Chami D.E. and Moujabber, M.E., 2016. Drought, climate change and sustainability of water in agriculture: A roadmap towards the NWRS2. *South African Journal of Science*, **112**(9-10), 1-4.

Chaussonnet, G., Riber, E., Vermorel, O., Cuenot, B., Gepperth, S. and Koch, R., 2013. Large Eddy Simulation of a prefilming airblast atomizer. *ILASS*.

Cochrane, J.H., 2005. Time series for macroeconomics and finance. *Manuscript, University of Chicago*, **15**, 16.

Craig, I. and Hancock, N., 2004. Methods for assessing dam evaporation-an introductory paper. In *Irrigation Australia 2004: Irrigation Association of Australia National Conference and Exhibition: Proceedings*. Irrigation Australia Ltd..

Dabral P.P. and Murry, M.Z., 2017. Modelling and forecasting of rainfall time series using SARIMA. *Environmental Processes*,**4**(2), 399-419.

Dabral, P.P., Mor, N. and Jhajharia, D., 2018. Time series modelling of monthly reference evapotranspiration for Bikaner, Rajasthan (India). Indian Journal of Soil Conservation, **46**(1), 42-51.

Dawood K.A., Rashid F.L. and Hashim A., 2013. Reduced evaporation losses from water reservoirs. *International Journal Energy and Environmental Research*, **1**(1), 23-29.

Delima, A.J.P., 2019. Application of Time Series Analysis in Projecting Philippines' Electric Consumption. *International Journal of Machine Learning and Computing*, **9**(5), 694-699.

Diamond R.E. and Jack, S., 2018. Evaporation and abstraction determined from stable isotopes during normal flow on the Gariep River, South Africa. *Journal of Hydrology*, **559**, 569-584.

Dwivedi, D.K., Sharma, G.R. and Wandre, S.S., 2017. Forecasting mean temperature using SARIMA Model for Junagadh City of Gujarat. *IJASR*, **7**(4), 183-194.

El-Ghonemy M.K, 2012.Water desalination systems powered by renewable energy sources: review. *Renewable Sustainable Energy Review*; **16**(3), 1537–1556.

El-Mallah, E.S. and Elsharkawy, S.G., 2016. Time-series modelling and short term prediction of annual temperature trend on Coast Libya using the box-Jenkins ARIMA Model. *Advances in Research*, 1-11.

Everitt, B.S. and Howell, D.C., 2021. *Encyclopedia of Statistics in Behavioral Science*, **Volume 2**. John Wiley & Sons, Ltd.

Fathian F., Fard A.F., Ouarda T.B., Dinpashoh Y. and Nadoushani S.M., 2019. Modelling streamflow time series using nonlinear SETAR-GARCH models. *Journal of Hydrology.*

Fithian, F., 2019. Dynamic memory of Urmia Lake water-level fluctuations in hydroclimatic variables. *Theoretical and Applied Climatology*, **138**(1), 591-603.

Francq, C., Horvath, L. and Zakoïan, J.M., Burns, P., 2011. Robustness of the Ljung-Box test and its rank equivalent. Available at SSRN 443560.Merits and drawbacks of variance targeting in GARCH models. *Journal of Financial Econometrics*, **9**(4), 619-656.

Ghasemi, A. and Zahediasl, S., 2012. Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*, **10**(2), 486.

Gleick P.H., 2014. Water, drought, climate change, and conflict in Syria. *Weather, Climate, and Society*, **6**(3), 331-340.

Gorjian S. and Ghobadian B., 2015. Solar desalination: A sustainable solution to water crisis in Iran. *Renewable and Sustainable Energy Reviews*, **48**, 571-584.

Gwate, O., 2018. Modelling plant water use of the grassland and thicket biomes in the Eastern Cape, South Africa: Towards an improved understanding of the impact of invasive alien plants on soil chemistry, biomass production and evapotranspiration. Faculty of Science, Institute for Water Research. Rhodes University.

Hassan, M., 2013. Evaporation estimation for Lake Nasser based on remote sensing technology. *Ain Shams engineering journal*, **4**(4), 593-604.

Hassan, A., Ismail, S.S., Elmoustafa, A. and Khalaf, S., 2018. Evaluating evaporation rate from high Aswan Dam Reservoir using RS and GIS techniques. *The Egyptian Journal of Remote Sensing and Space Science*, **21**(3), 285-293.

Hazelton, M.L., 2011. Methods of Moments Estimation BT–International Encyclopedia of Statistical Science, edited by: Lovric, M. 816-817

Hensley, M., le Roux, P.A., Botha, J.J. and van Rensburg, L.D., 2019. The role of water conservation strategies and benchmark ecotopes for increasing yields in South Africa's semi-arid croplands. *Water SA*, **45**(3), 393-399.

Hipel, K.W. and McLeod, A.I., 1994. *Time series modelling of water resources and environmental systems*. Elsevier.

Hughes D.A., 2019. A simple approach to estimating channel transmission losses in large South African river basins.*Journal of Hydrology: Regional Studies*, **25**, 100619.

Hurlin, C., 2013. *Maximum likelihood estimation. Advanced econometrics*. The University of Orleans.

Ihaka, R., 2005. *Time Series Analysis*. Lecture Notes for, **475**.

Issaka M., 2015. *Modelling the dynamic relationship between rainfall and temperature in Kassena-Nankana municipality*.

Jovanovic, N.Z., Jarmain, C., De Clercq, W.P., Vermeulen, T. and Fey, M.V., 2011. Total evaporation estimates from a Renosterveld and dry-land wheat/fallow surface at the Voëlvlei Nature Reserve (South Africa). *Water SA*, **37**(4), 471-482.

Khair, U., Fahmi, H., Al Hakim, S. and Rahim, R., 2017, December. Forecasting error calculation with mean absolute deviation and mean absolute percentage error. In *Journal of Physics: Conference Series* (**Vol. 930**, No. 1, 012002). IOP Publishing.

Kuun G.F., 2009. The construction of dams to ensure water security in South Africa.

Lai, C.T.T., 2017. *Respiratory disease diagnosis for dolphin using breath data*, Doctoral dissertation, Texas A&M University-Corpus Christi.

Lancsar, E., Fiebig, D.G. and Hole, A.R., 2017. Discrete choice experiments: a guide to model specification, estimation and software. *Pharmacoeconomics*, **35**(7), 697-716.

Machekposhti K.H., Sedghi H., Telvari A. and Babazadeh H., 2018. Modelling climate variables of rivers basin using time series analysis (case study: Karkheh River basin at Iran). *Civil Engineering Journal*, **4**(1), 78-92.

]Machete, M. and Shadung, J.M., 2019. Detection of selected agricultural pesticides in river and tap water in Letsitele, Lomati and Vals–Renoster catchments, South Africa. *Water SA*, **45**(4), 716-720.

Makungo, R. and Odiyo, J.O., 2017. Estimating groundwater levels using system identification models in Nzhelele and Luvuvhu areas, Limpopo Province, South Africa. *Physics and Chemistry of the Earth*, Parts A/B/C, **100**, 44-50.

Manuca, R. and Savit, R., 1996. Stationarity and nonstationarity in time series analysis. *Physica D: Nonlinear Phenomena*, **99**(2-3), 134-161.

Mapholi M., 2018. *Facilitating an improved Integrated Water Resources Management (IWRM) model in Madibeng Local Municipality* ,Doctoral dissertation, North-West University.

Maponya P.I., 2012.*Climate change and agricultural production in Limpopo Province: impacts and adaptation options*,Doctoral dissertation.

Martin, J., De Adana, D.D.R. and Asuero, A.G., 2017. Fitting models to data: residual analysis, a primer. Uncertainty quantification and model calibration, 133. Martin, J., De Adana, D.D.R. and Asuero, A.G., 2017. Fitting models to data: residual analysis, a primer. *Uncertainty quantification and model calibration*, 133.

Martinez-Granadose D., Maestre-Valero J.F., Calatrava J. and Martínez-Alvarez V.,2011. The economic impact of water evaporation losses from water reservoirs in the Segura basin, SE Spain. *Water Resources Management*, **25**(13), 3153.

Matringe, O. and Guida, T., 2004. Application of GARCH Models in Forecasting the Volatility of Agricultural Commodities.

Masupha, T.E. and Moeletsi, M.E., 2018. Analysis of potential future droughts limiting maize production, in the Luvuvhu River catchment area, South Africa. *Physics and Chemistry of the Earth*, Parts A/B/C, **105**, 44-51.

Mathivha, F., Sigauke, C., Chikoore, H. and Odiyo, J., 2020. Short-Term and Medium-Term Drought Forecasting Using Generalized Additive Models. *Sustainability*, **12**(10), 4006.

McKenzie, R.S. and Craig, A.R., 2001. Evaluation of river losses from the Orange River using hydraulic modelling. *Journal of Hydrology*, **241**(1-2), 62-69

Meissner R., Funke N., Nortje K., Jacobs-Mata I., Moyo E., Steyn M., Shadung J., Masangane W. and Nohayi N., 2018. Water security at local government level in South Africa: A qualitative interview-based analysis. *The Lancet Planetary Health*, **2**, S17.

Mekkonen, M. and Hoekstra, A. (2010) A Global and High-Resolution Assessment of the Green, Blue and Grey Water Footprint of Wheat. *Hydrology and Earth Systems Sciences*, **14**, 1259-1276.

Mohanasundaram, S., Narasimhan, B. and Kumar, G.S., 2013, The Significance of Autocorrelation and Partial Autocorrelation on Univariate Groundwater Level Rise (Recharge) Time Series Modelling. Published by Association of Global Groundwater Scienists, India, **Vol.2**(1), 2321- 4783.

Montgomery, M.R., 2008. The urban transformation of the developing world. *Science*, **319**(5864), pp.761-764.

Montgomery, D.C., Jennings, C.L. and Kulahci, M., 2015. *Introduction to time series analysis and forecasting*. John Wiley & Sons.

Mosase, E. and Ahiablame, L., 2018. Rainfall and temperature in the Limpopo River Basin, Southern Africa: means, variations, and trends from 1979 to 2013. *Water SA*, **10**(4), 364.

Mosase E., Ahiablame, L. and Srinivasan, R., 2019. Spatial and temporal distribution of blue water in the Limpopo River Basin, Southern Africa: A case study.*Ecohydrology & Hydrobiology*, **19**(2), 252-265.

Motoshita M., Ono Y., Pfister S., Boulay A.M., Berger M., Nansai K., Tahara K., Itsubo N. and Inaba A., 2018. Consistent characterisation factors at midpoint and endpoint relevant to agricultural water scarcity arising from freshwater consumption. *The International Journal of Life Cycle Assessment*, **23**(12), 2276-2287.

Musetha M.A., 2016.*The impact of climate change on agricultural crop production in the Vhembe District Municipality, Limpopo Province South Africa*. Doctoral dissertation.

Mushtaq, R., 2011. Augmented dickey fuller test.

Mzezewa J., Misi, T. and Van Rensburg, L., 2010. Characterisation of rainfall at a semi-arid ecotope in the Limpopo Province (South Africa) and its implications for sustainable crop production. *Water SA*, **36**(1)

Naabil E., Lamptey B.L., Arnault J., Olufayo A. and Kunstmann H., 2017. Water resources management using the WRF-Hydro modelling system: Case-study of the Tono dam in West Africa. *Journal of Hydrology: Regional Studies*, **12**, 196-209.

Nau, R., 2014. Forecasting with moving averages. Fuqua School of Business, Duke University, 1-3.

Nhemachena C. and Hassan, R., 2007. *Micro-level analysis of farmers adaption to climate change in Southern Africa*. International Food Policy Research Institute.

Ostertagova, E. and Ostertag, O., 2011, September. *The simple exponential smoothing model*. In The 4th International Conference on Modelling of Mechanical and Mechatronic Systems, Technical University of Košice, Slovak Republic, Proceedings of conference, 380-384.

Perera, I. and Silvapulle, M.J., 2018. Specification tests for time series models with GARCH-type conditional variance. Available at SSRN 3141822.

Perry, M.B., 2010. *The exponentially weighted moving average* . Wiley Encyclopedia of Operations Research and Management Science.

Qasem S.N., Samadianfard S., Kheshtgar, S., Jarhan S., Kisi O., Shamshirband S. and Chau K.W., 2019. Modelling monthly pan evaporation using wavelet support vector regression and wavelet artificial neural networks in arid and humid climates. *Engineering Applications of Computational Fluid Mechanics*, **13**(1), 177-187.

Reinert, G., 2010. *Time series*. Hillary Term, 12.

Salem A., Dezső, J. and El-Rawy, M., 2019. Assessment of groundwater recharge, evaporation, and runoff in the Drava Basin in Hungary with the WetSpass Model. *Hydrology*, **6**(1), 23.

Sebbar, A., Heddam, S., Kisi, O., Djemili, L. and Houichi, L., 2020. Comparison of Evolving Connectionist Systems (ECoS) and Neural Networks for Modelling Daily Pan Evaporation from Algerian Dam Reservoirs. *Water Resources in Algeria-Part I: Assessment of Surface and Groundwater Resources*, 161-179.

Shabalala Z.P., Moeletsi M.E., Tongwane M.I. and Mazibuk, S.M., 2019. Evaluation of Infilling Methods for Time Series of Daily Temperature Data: Case Study of Limpopo Province, South Africa. *Climate*, **7**(7), 86.

Shahidi, A., Ramezani, Y., Nazeri-Tahroudi, M. and Mohammadi, S., 2020. Application of vector autoregressive models to estimate pan evaporation values at the Salt Lake Basin, Iran. *Idojaras/Quartely journal of the hungarian meterological service*, **124**(4), 463-482.

Shumway, R.H., Stoffer, D.S. and Stoffer, D.S., 2000. *Time series analysis and its applications* **(Vol. 3)**. New York: springer.

Stehlikova, B., 2005. Modelling volatility clusters with application to two-factor interest rate models. J. Electr. Engrg, **56**, 12.

Stine, R.A., 2017. Explaining normal quantile-quantile plots through animation: the water-filling analogy. The American Statistician, **71**(2), 145-147.

Sule B.F. and Ajala, O.F., 2017. Long-term Estimates of Reservoir Evaporation Using ARIMA Model and Impact on Water Supply: A Case Study of Erinle Dam, Osun State, Nigeria. *International Journal of Sciences*, **6**(09), 29-38.

Sun H., Yan D., Chen H., Zhou, J. and Zhang Y., 2015. Modelling heteroscedasticity of reference evapotranspiration series with GARCH family models. *Transactions of the Chinese Society of Agricultural Engineering*, **31**(7), 131-136.

Swanson, D.A., Tayman, J. and Bryan, T.M., 2011. MAPE-R: a rescaled measure of accuracy for cross-sectional subnational population forecasts. *Journal of Population Research*, **28**(2), 225-243.

Taheri Tizro A., Ghashghaie, M., Georgiou, P. and Voudouris, K., 2014. Time series analysis of water quality parameters.*Journal of Applied Research in Water and Wastewater*, (1),40-50.

Tezel, G. and Buyukyildiz, M., 2016. Monthly evaporation forecasting using artificial neural networks and support vector machines. *Theoretical and applied climatology*, 124(1-2), 69-80.

Vandekerckhove, J., Matzke, D. and Wagenmakers, E.J., 2014. *Model comparison and the principle of parsimony* (1-29). eScholarship, University of California.

Vanderkelen, I., Van Lipzig, N.P. and Thiery, W., 2018. Modelling the water balance of Lake Victoria (East Africa)-Part 1: Observational analysis. *Hydrology and Earth System Sciences*, **22**(10),

Velicer, W.F. and Velicer, J.L., 2003. Time series analysis. *Research methods in psychology*, 2.

Wine M.L., Rimmer A. and Laronne J.B., 2019. Agriculture, diversions, and drought shrinking Galilee Sea. *Science of The Total Environment*, **651**, 70-83.

Yue, C., Wang, B. and Zhu, B., 2018. Thermal analysis for the evaporation concentrating process with high boiling point elevation based exhaust waste heat recovery. *Desalination*, **436**, 39-47.

Zhang, S., 2016. Adaptive spectral estimation for nonstationary multivariate time series. *Computational Statistics & Data Analysis*, **103**, 330-349.

Zhao, P., Yang, Y. and Zhou, X., 2021. Weighted empirical likelihood inferences for a class of varying coefficient ARCH-M models. *Journal of Nonparametric Statistics*, **33**(1), 1-20.

Zhao, G. and Gao, H., 2019. Estimating reservoir evaporation losses for the United States: Fusing remote sensing and modelling approaches. *Remote Sensing of Environment*, **226**, 109-124. 5509-5525.

Ziervogel G., Taylor, A., Thomalla, F., Takama, T. and Quinn, C., 2006. *Adapting to climate, water and health stresses: insights from Sekhukhune, South Africa*. Stockholm Environment Institute (SEI).

Appendix

Tables and plots for Chapter 4

Table 1 Descriptive statistics

Station names	Latitude	Longitude	Altitude
Lephalale MolokoDam	-23,76571	27,74091	826
Sekhukhune Ga-RanthoDam	-24,87423	29,96795	1152
Sekhukhune Leeukraal DeHoopDam	-24,491585	29,83511	1446
Thohoyando LephepheDam	-22,73461	30,52188	550

Table 2 ARCH model of water evaporation for Mokolo dam

Coefficients			
	AR1	MA1	MA2
	0.5748	-1.3619	0.3973
s.e.	0.1575	0.1708	0.1550

Table 3 ARCH model of water evaporation for Ga-Rantho dam

Coefficients			
	AR1	MA1	MA2
	0.3571	-1.1564	0.2115
s.e.	0.1193	0.1233	0.1120

Table 4 ARCH model of water evaporation for Leeukraal DeHoop dam

Coefficients				
	AR1	MA1	MA2	MA3
	-0.0304	-0.4305	-0.1688	-0.0692
s.e.	0.2821	0.2816	0.1342	0.0542

Table 5ARCH model of water evaporation for Luphephe dam

Coefficients				
	AR1	MA1	MA2	MA3
	2.289946	0.498327	-0.020096	0.197179
s.e.	1.306119	0.566913	0.427403	0.365485

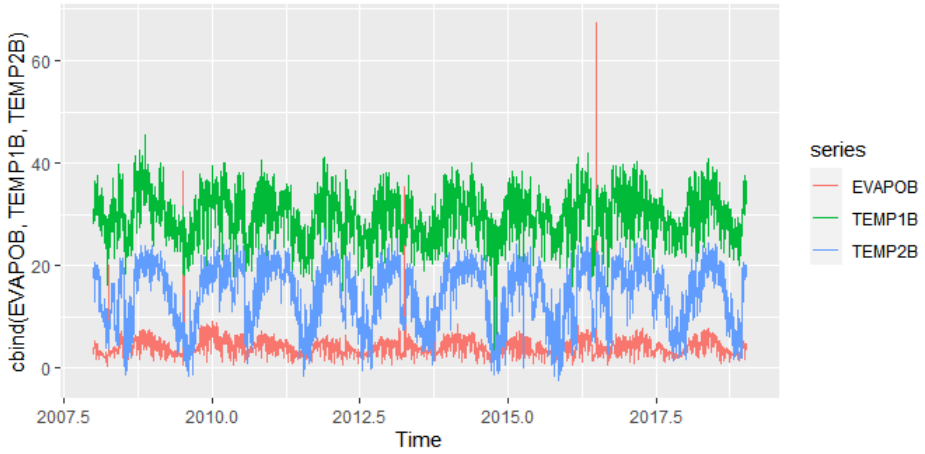


Figure 1 VAR series plot of evaporation and temperature Mokolo dam

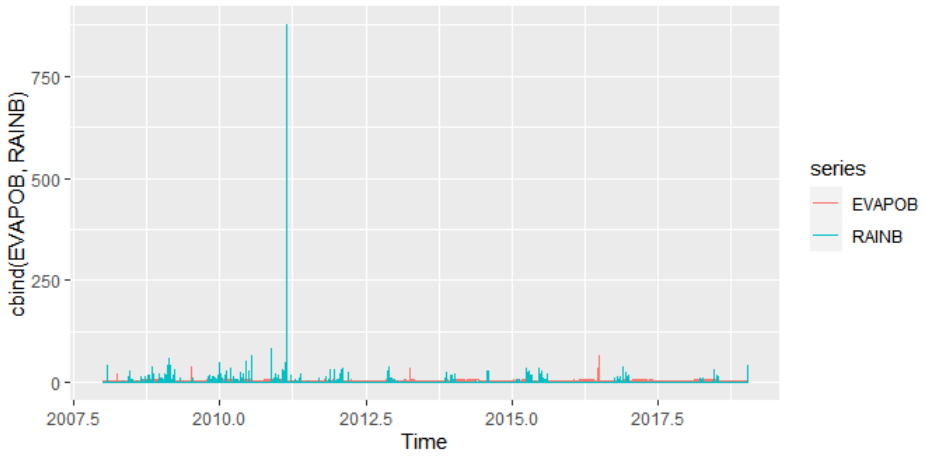


Figure 2 VAR series plot of evaporation and rain Mokolo dam

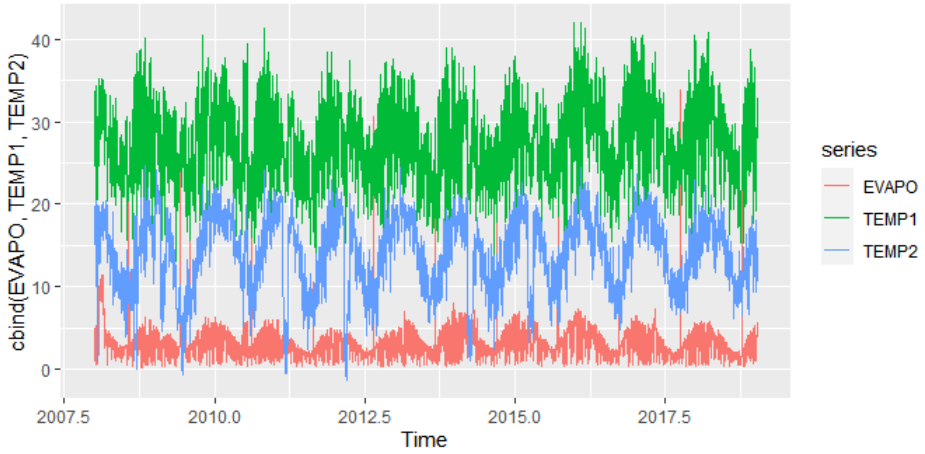


Figure 3 VAR series plot of evaporation and temperature Ga-Rantho dam

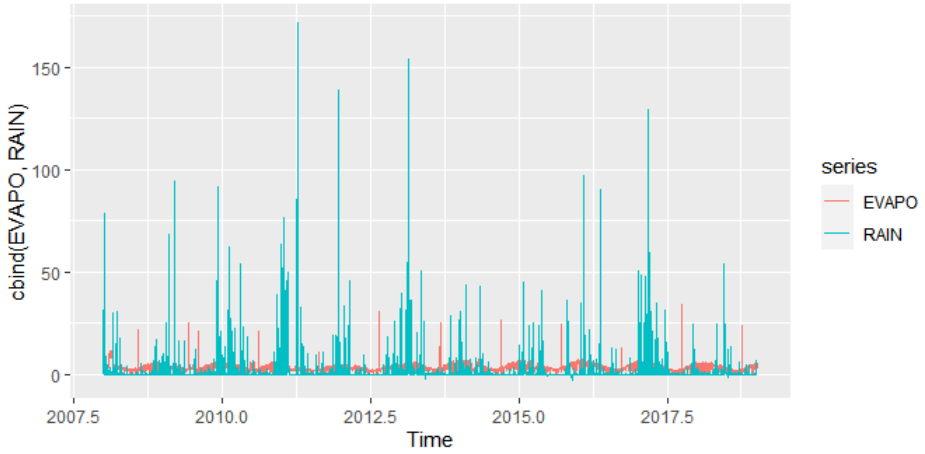


Figure 4 VAR series plot of evaporation and rain Ga-Rantho dam

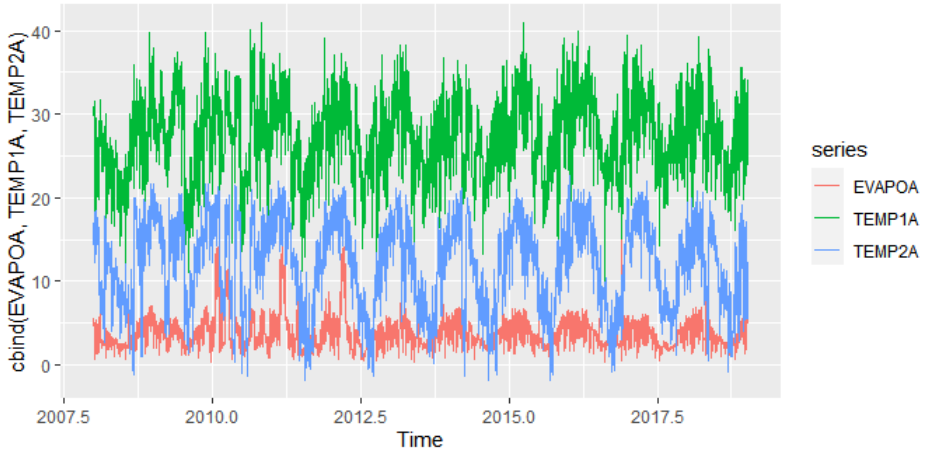


Figure 5 VAR series plot of evaporation and temperature Leeukraal DeHoop dam

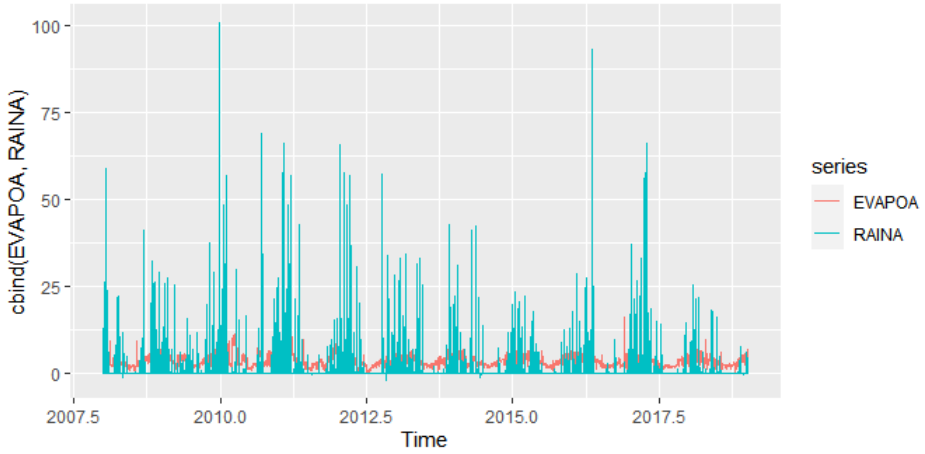


Figure 6 VAR series plot of evaporation and rain Leeukraal DeHoop dam

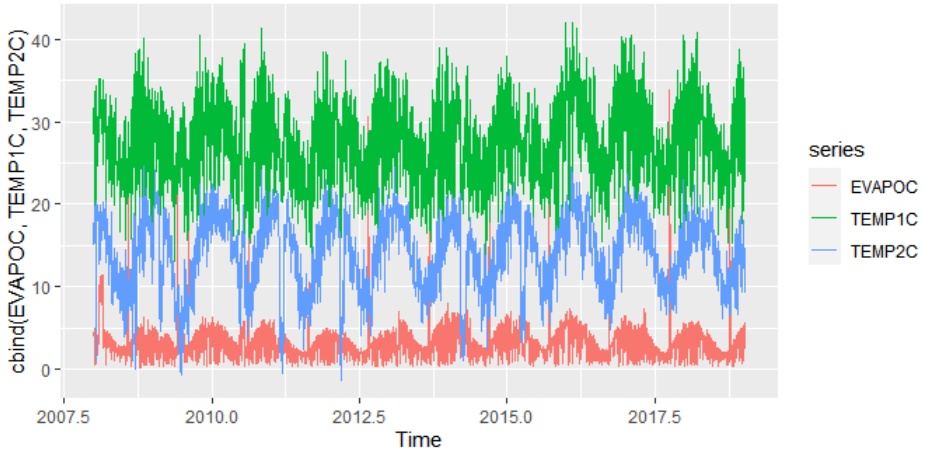


Figure 7 VAR series plot of evaporation and temperature Luphephe dam

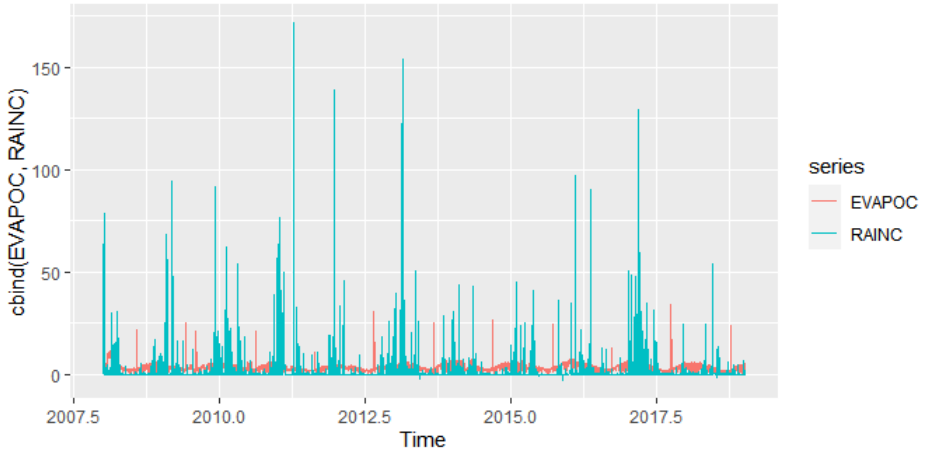


Figure 8 VAR series plot of evaporation and rain Luphephe dam

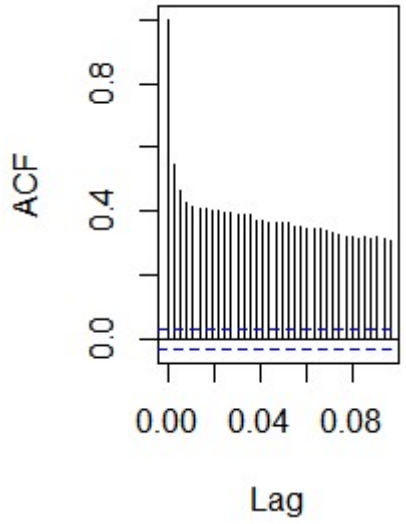


Figure 9 GARCH ACF evaporation rate Mokolo dam

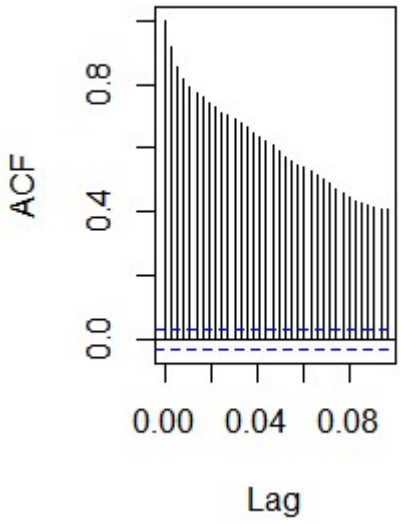


Figure 10 GARCH ACF temperature Mokolo dam

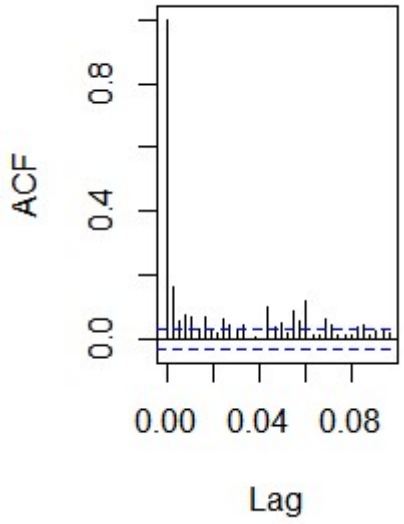


Figure 11 GARCH ACF rainfall Mokolo dam

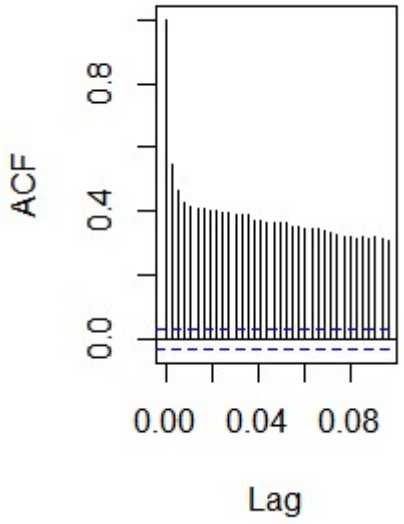


Figure 12 GARCH ACF evaporation rate Ga-Rantho dam

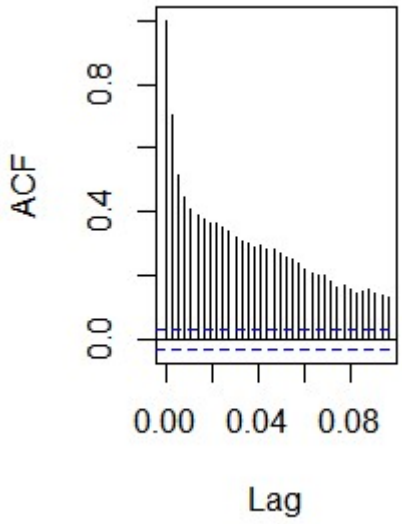


Figure 13 GARCH ACF temperature Ga-Rantho dam

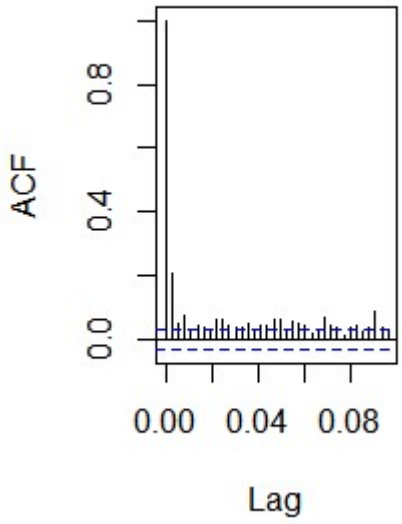


Figure 14 GARCH ACF rainfall Ga-Rantho dam

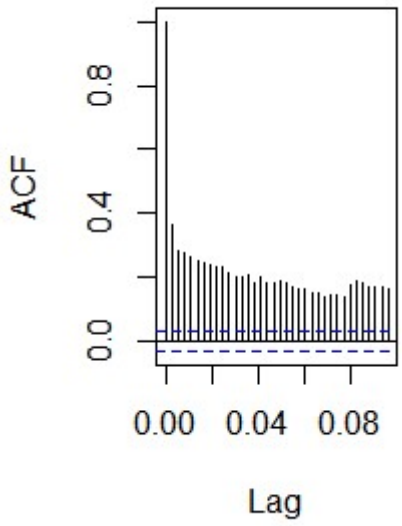


Figure 15 GARCH ACF evaporation rate Leeukraal DeHoop dam

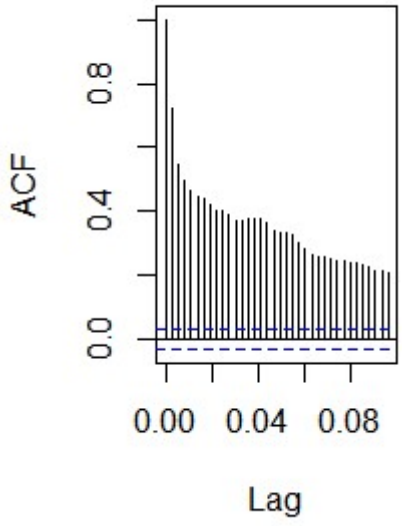


Figure 16 GARCH ACF temperature Leeukraal DeHoop dam

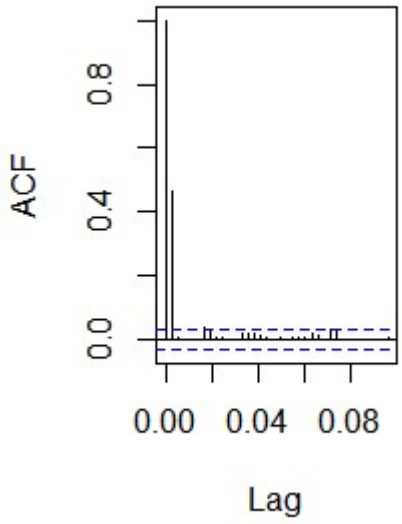


Figure 17 GARCH ACF rainfall Leeukraal DeHoop dam

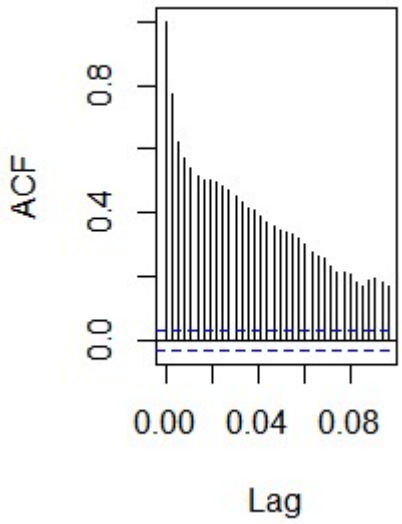


Figure 15 GARCH ACF evaporation rate Luphephe dam

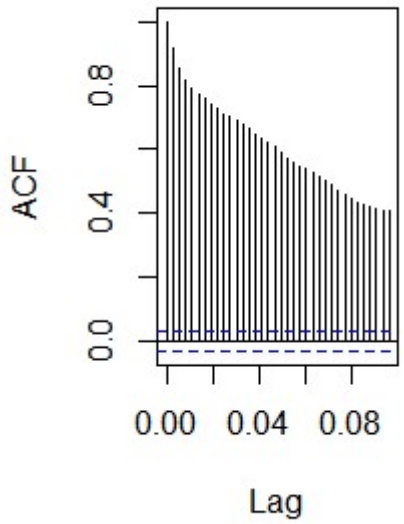


Figure 16 GARCH ACF temperature Luphephe dam

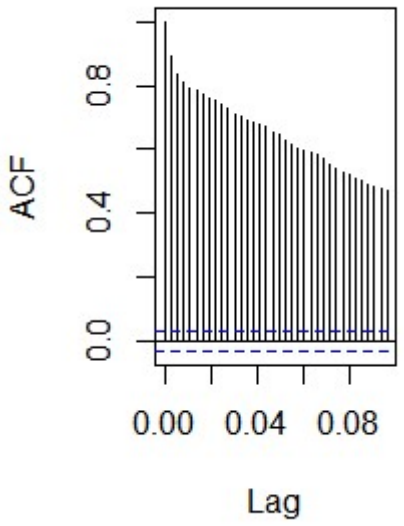


Figure 17 GARCH ACF rainfall Luphephe dam

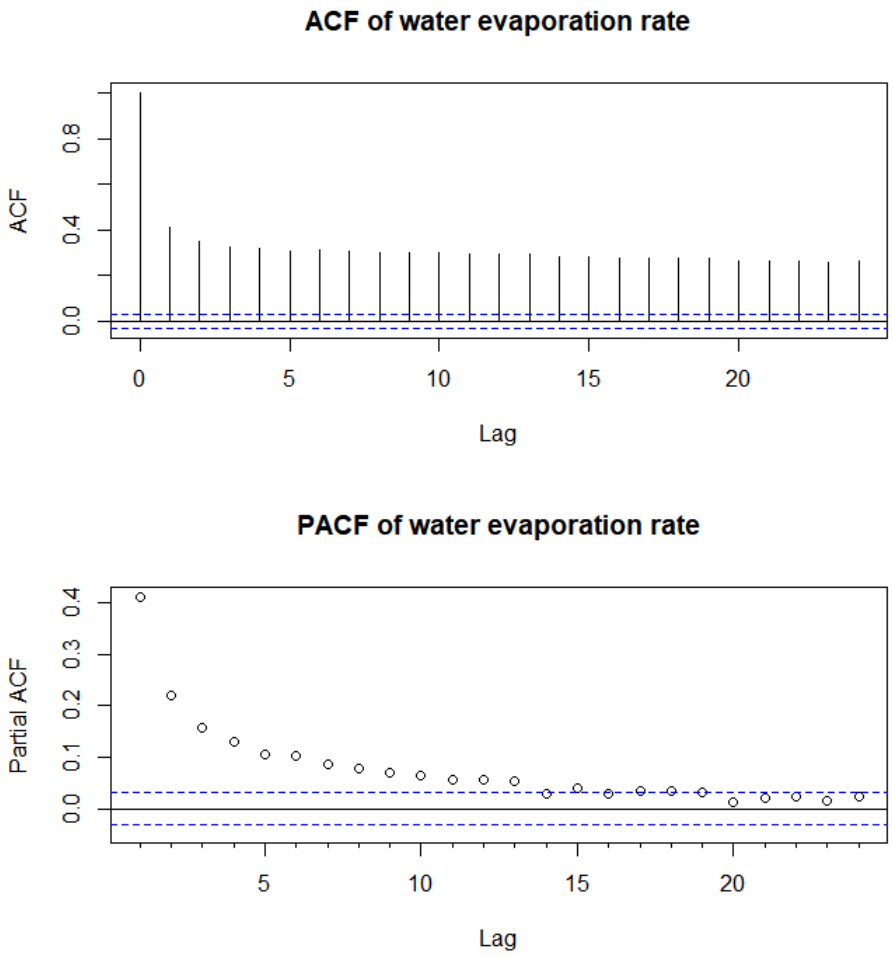
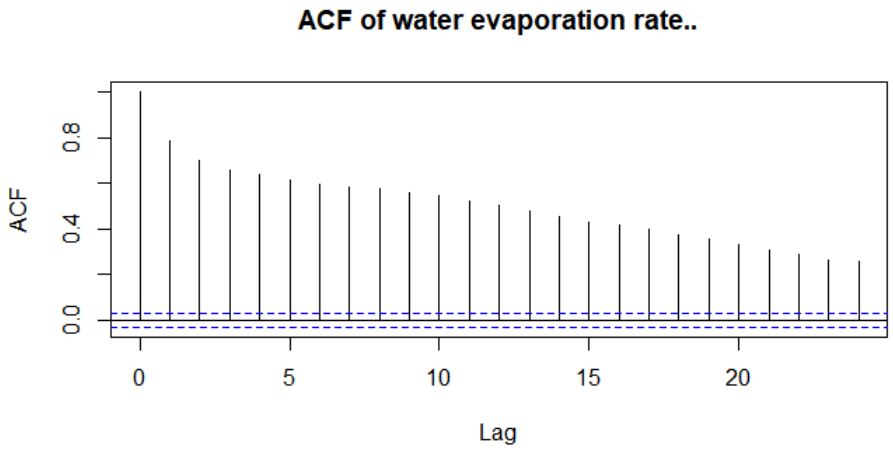


Figure 18 ARCH ACF and PACF evaporation rate Mokolo dam



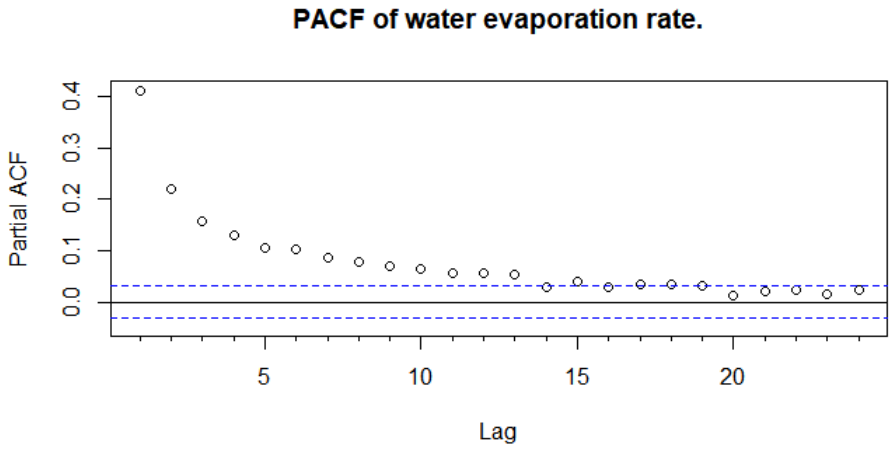


Figure 20 ARCH ACF and PACF evaporation rate Ga-Rantho dam

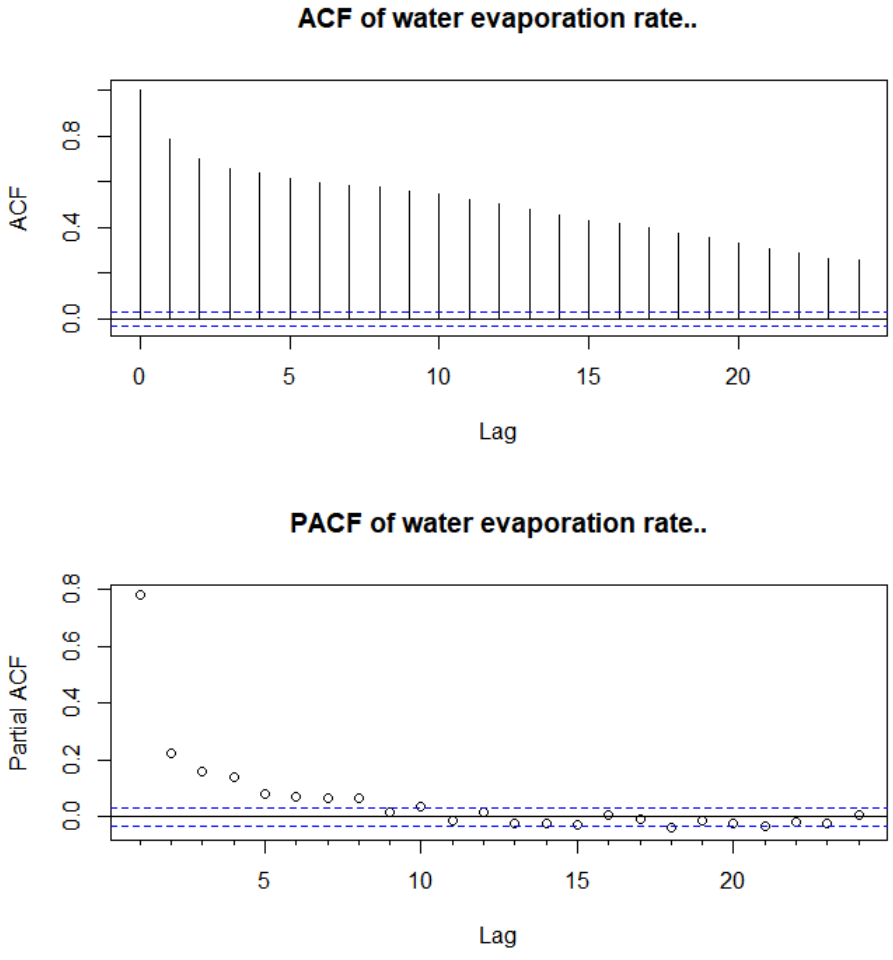


Figure 21 ARCH ACF and PACF evaporation rate Leeukraal DeHoop dam

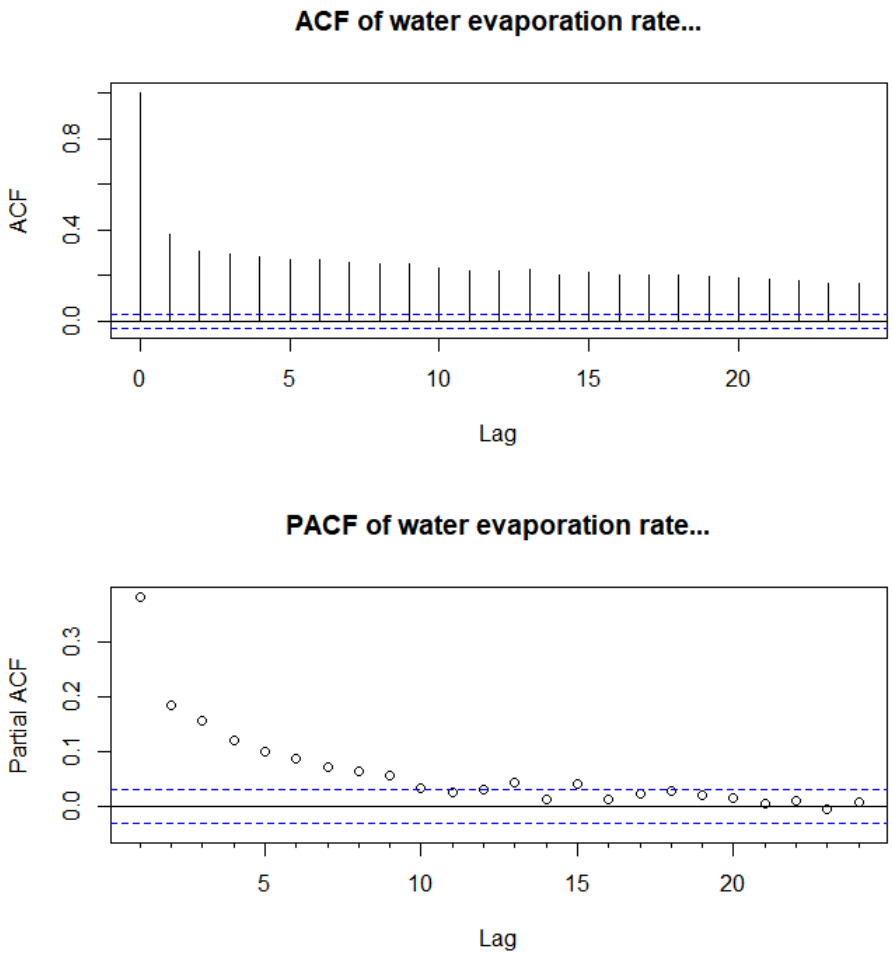


Figure 22 ARCH ACF and PACF evaporation Luphephe dam

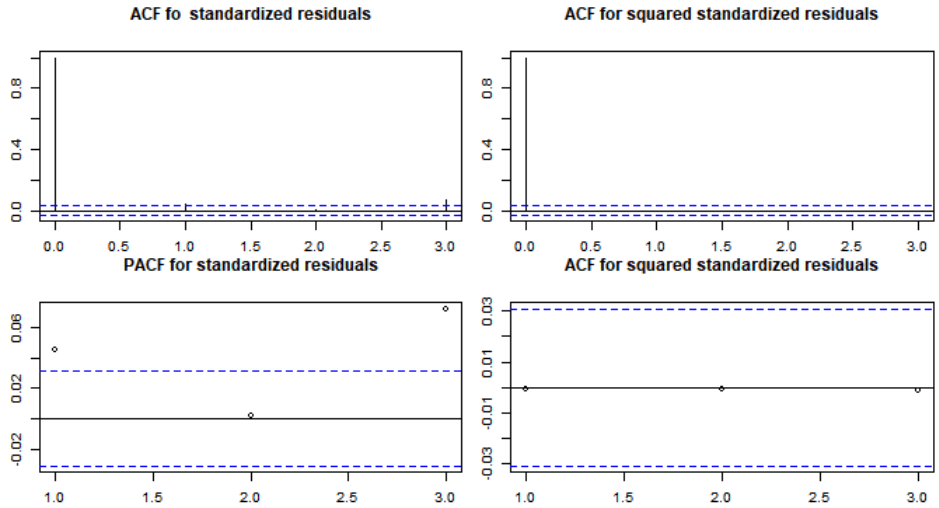


Figure 23 Residuals for evaporation rate Mokolo dam

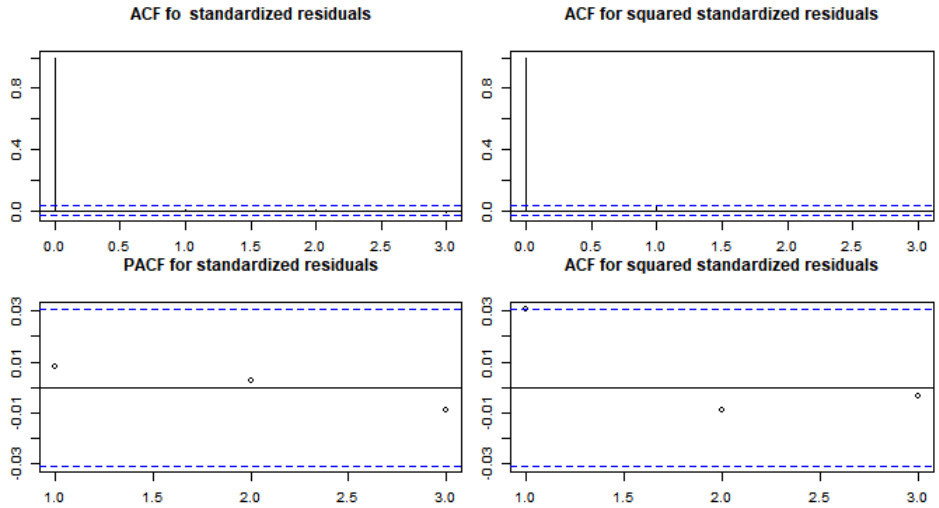


Figure 24 Residuals for evaporation rate Ga-Ranth dam

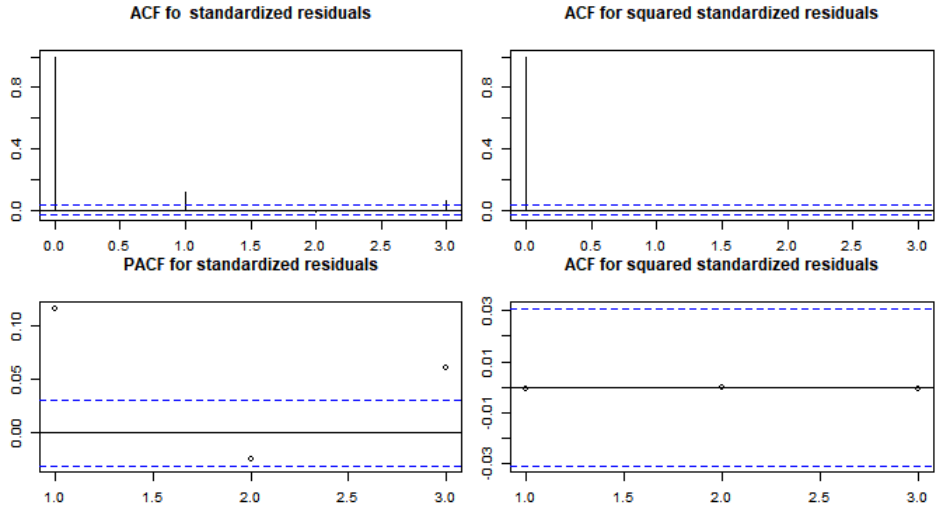


Figure 25 Rsiduals for evaporation rate Leeukraal DeHoop dam