PREDICTION OF EVAPOTRANSPIRATION FOR OPTIMUM WATER USE IN AGRICULTURE: A CASE STUDY OF KEISKAMMA IRRIGATION SCHEME, EASTERN CAPE, SOUTH AFRICA

MBULELO PHESA

Dissertation submitted in fulfilment of the requirements for the Degree

MASTER OF ENGINEERING IN CIVIL ENGINEERING

in the
Department of Civil Engineering
Faculty of Engineering, Built Environment and
Information Technology
at the
Central University of Technology, Free State

Bloemfontein

June 2022

Supervisor: Prof. Yali E. Woyessa (PhD)

Co-supervisor: Dr N. Mbatha (PhD)

DECLARATION OF INDEPENDENT WORK

I, MBULELO PHESA, identity number ______________ and student number __________, do hereby declare that this dissertation submitted to the Central University of Technology, Free State, for the Master of Engineering in Civil Engineering is my own original research work; and complies with the code of Academic Integrity, as well as other relevant policies, procedures, rules and regulations of the Central University of Technology, Free State; and has not been submitted before to any institution by myself or any other person in fulfilment of the requirements for attainment of any qualification.

__________________________________   **18/11/2022**

SIGNATURE OF STUDENT   **DATE**

DEDICATION

I dedicate this work to my late parents, Mr Mkalwa Phesa and Mrs Matitiya Phesa, who believed and loved education even though they never went to school because of the circumstances of their time.

ACKNOWLEDGEMENT

I extend my gratitude to the following who played a key role to the success of my study:

- God Almighty for keeping me under his mighty arm throughout the study.
- Professor Yali Woyessa, my supervisor, for his professional contribution, support and guidance all the time towards the success of this study. Thank you for the encouragement and guidance you gave and for believing in me.
- Dr N. Mbatha, my co-supervisor, for supporting me throughout the study. He shared tirelessly all the information he has on modelling, writing and his expertise in ARIMA, ANN and Hybrid Model. Thank you for being the tireless coach.
- Dr A. Ikudayisi, for keeping on motivating me and also guiding me as the WSU appointed mentor under nGAP to oversee my study. Thank you for your continued support and tolerance.
- Dr T. Mandindi, for her wise words of encouragement and for playing a motherly role by encouraging me all the time.
- The entire Walter Sisulu University Civil Engineering staff for your support and indefatigable assistance.
- My family (brothers, sisters and my children) and friends for keeping on encouraging me throughout the course of the study.
- Department of Higher Education and Training for funding the study through the nGAP Programme managed by Walter Sisulu University Centre of Learning and Development.
- MerSETA for also funding the project through the WSU/MerSETA 2020-2023 Project.

ABSTRACT

Evapo-transpiration (ET) is one of the crucial elements of the hydrological cycle which expedites constant precipitation through the process of condensation. The accurate prediction of ET is essential for irrigated agriculture as it informs appropriate planning and contributes positively to the daily supervision of the irrigation scheme. However, because of the limited data in arid and semi-arid regions, which have been used widely in the traditional Penman-Monteith approach, alternative, reliable and more powerful techniques are used to predict ET. South Africa is one of many countries that fall under the semi-arid zones where the degree of evapo-transpiration is more than the rainfall rate. The aim of this study was to predict evapo-transpiration in the Keiskammahoek Irrigation Scheme located in Eastern Cape, South Africa, using three, time series, prediction models, namely, Auto Regressive Integrated Moving Average (ARIMA), Artificial Neural Networks (ANNs) and Hybrid (ARIMA-ANNs). ARIMA and ANNs models have been used mostly over the years to predict the linear and non-linear time series, and the Hybrid model, developed by Zhang was also applied because of its ability to capture both the linear and non-linear time series. The 18 years (2001 to 2018) ET time series data was extracted from Google Earth Engine, using java script, at Keiskammahoek Irrigation Scheme. Prior to the prediction of ET time series, the time series data were analysed to understand the behaviour of the time series. A detailed time analysis of Keiskamma River Streamflow and Sandile Dam monthly volume, which are water supply sources close to the study area, were analysed, using time series analysis methods such as: the Break for Additive Seasonal (BFAS) and Trend, Wavelength Analysis, Wavelet Coherence, Correlation Statistics, Theil-Sen plots, Man-Kendall Test, Sequential Mann-Kendall Test and Multi-Linear Regression Analysis. Furthermore, tele-connection analysis between the satellite-derived et time series for the study area and other parameters, such as the Normalised Difference Vegetation Index (NDVI), Normalised Difference Water Index (NDWI), Normalised Difference Drought Index (NDDI), and Precipitation(P), was performed. Through use of the Mann-Kendall Trend Test, it was noted that et has increased over the years with the z-score reaching +3.898 which is greater than +1.96 which indicates the significance of the trend, in contrast to the z-score for precipitation which is equal to -2.6134, indicating a significant decrease in P in the study area over the 18-year period.

This trend, and its significance, were noted also using the Sequential Mann-Kendall method. Using the Multi Linear Regression, a statistically significant relationship was also noted between ET, with p-value < $2X10^{-16}$ and NDVI, p-value equal to 7.89×10^{-11}, for Stream Flow p-value equal to $2.32X10^{-06}$, for P p-value equal to < $2Ex10^{-16}$ and for NDDI p-value equal to 0.0208. The significant relationship between these variables is indicated by a p-value less than 0.05.. Using ARIMA, ANNs and Hybrid(ARIMA-ANNs), the ET at Keiskammahoek Irrigation Scheme was predicted successfully for 3 years (2015 to 2018). Furthermore, the three models were combined to assess the quality of prediction further and ET was once more predicted successfully. To select the best performing model for prediction of ET, the predicted results for three applied modelling techniques were evaluated using four, well accepted, model performance statistics, namely, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Pearson's Correlation Coefficient (R). The results of this study shows that, the hybrid (ARIMA-ANN) model outperformed both the ARIMA and ANN consecutively with less values of the statistical performance evaluation show-ing RMSE = 33.80, MAE = 27.02, MAPE = 17.31, and R = 0.94 compared to higher values of ARIMA and ANN In general, these forecasting results show the superiority of the Hybrid (ARIMA-ANN) model over ARIMA and ANN.

Keywords: Agriculture, evapo-transportation, irrigation water, prediction, planning and management, Auto Regressive Integrated Moving Average, Artificial Neural Networks

TABLE OF CONTENTS

DEDICATION .... **i**

ACKNOWLEDGEMENT .... **ii**

ABSTRACT .... **iii**

TABLE OF CONTENTS .... **v**

LIST OF FIGURES .... **ix**

LIST OF TABLES .... **xi**

ACRONYMS AND ABBREVIATIONS .... **xii**

CHAPTER 1: INTRODUCTION **1**

1.1 INTRODUCTION 1

1.2 PROBLEM STATEMENT 3

1.3 AIM 4

1.4 OBJECTIVES 4

1.5 SIGNIFICANCE OF THE STUDY 4

1.6 STRUCTURE OF THE DISSERTATION 5

CHAPTER 2: LITERATURE REVIEW **6**

2.1 STATE OF IRRIGATION IN SOUTH AFRICA 6

2.2 HYDROLOGICAL CYCLE 7

2.3 EVAPO-TRANSPIRATION 8

2.4 EVAPORATION 9

2.5 TRANSPIRATION 11

2.6 IRRIGATION SCHEDULING 11

2.7 STATE OF DROUGHT IN EASTERN CAPE 12

2.8 DROUGHT ASSESSMENT INDICES 15
2.8.1 NORMALISED DIFFERENCE WATER INDEX 15
2.8.2 NORMALISED DEFERENCE VEGETATION INDEX 16
2.8.3 NORMALISED DIFFERENCE DROUGHT INDEX 16
2.9 EVAPO-TRANSPIRATION PREDICTION TECHNIQUES 17
2.10 TIME-SERIES PREDICTION 18
2.10.1 Auto-regressive Integrated Moving Average (ARIMA) Models 19
2.10.2 Artificial Neural Networks (ANNs) 26
2.10.3 Hybrid (ARIMA-ANN) Model 29
2.10.3.1 Use of different hybrid models 29
2.10.3.2 Averaged models 32
2.11 R AND R-STUDIO 33
CHAPTER 3: MATERIAL AND METHODS 35
3.1 DESCRIPTION OF THE STUDY SITE 35
3.2 RESEARCH DESIGN 37
3.3 DATA COLLECTION 38
3.4 DATA ANALYSIS 41
3.4.1 WAVELET ANALYSIS 41
3.4.2 Analysis of Time Series Coherence Using Wavelet Coherence 42
3.4.3 Correlation Statistics Test Using Correlation Statistics 43
3.4.4 Trend Slope Assessment Using Theil-Sen’s Estimator 43
3.4.5 Trend Analysis Using Mann-Kendall Test 44
3.4.6 Sequential Mann-Kendall (SQ-MK) Test 46
3.4.7 Multi-Linear Regression (MLR) Model For Checking Connection Between Dependent Variables 47

3.4.8 Auto-Regressive Integrated Moving Average (ARIMA) Model 48
3.4.9 Auto-ARIMA 49
3.4.10 Artificial Neural Network (ANN) Model 50
3.4.11 Hybrid (ARIMA-ANN) Model 54
3.5 COMPARISON OF MODEL PERFORMANCE 56
3.5.1 Root Mean Square Error (RMSE) 56
3.5.2 Mean Absolute Percentage Error (MAPE) 57
3.5.3 Mean Absolute Error 57
3.5.4 Pearson's Correlation Coefficient... 58
CHAPTER 4: RESULTS AND DISCUSSION 59
4.1 INTRODUCTION 59
4.2 TIME SERIES ANALYSIS.. 59
4.2.1 The BFAST Analysis of the Time Series... 60
4.2.2 Wavelength Analysis.. 64
4.2.3 Wavelet Coherence 65
4.2.4 Correlation Statistics 66
4.2.5 Theil-Sen Plot 67
4.2.6 Man-Kendall Test... 70
4.2.7 Sequential Mann-Kendall (SQ-MK) Test.. 71
4.2.8 Multi Linear Regression Analysis... 73
4.3 MODEL SELECTION AND DISCUSSION OF RESULTS........................ 74
4.3.1 ARIMA Model Training and Validation 75
4.3.2 ARIMA Prediction... 78
4.3.3 ANN Models Training and Prediction 79

4.3.4 Hybrid (ARIMA-ANN) Model Prediction 81

4.3.5 Averaged (ARIMA, ANN and Hybrid (ARIMA-ANN) Model Prediction 83

4.3.6 Comparison of Models 85

4.4 PREDICTING FUTURE ET USING THE HYBRID (ARIMA-ANN) MODELS 86

CHAPTER 5: CONCLUSION AND RECOMMENDATIONS 88

5.1 SUMMARY 88

5.2 CONCLUSION 90

5.3 RECOMMENDATIONS 90

REFERENCES 91

APPENDIX 106

LIST OF FIGURES

Figure 2.1: Main reservoirs and fluxes of water 8

Figure 3.1: Keiskammahoek Irrigation Scheme 35

Figure 3.2: ALHF and precipitation of Keiskammahoek Irrigation Scheme. 36

Figure 3.3: A flow chart of the research design. 37

Figure 3.4: Extracted Keiskammahoek Irrigation Scheme 39

Figure 3.5: Sandile Dam 40

Figure 3.6: Taxonomy of feed-forward networks 51

Figure 3.7: Taxonomy of recurrent/ feedback networks 52

Figure 3.8: Typical structure of ANN 54

Figure 4.1**:** Time series for ET, P, SF, NDVI, NDWI, NDDI and MV (2001-2018) 60

Figure 4.2: BFAST graphs for ET (a), NDWI (b), P (c), NDVI (d), Dam MV (e) and SF (f) time series components 62

Figure 4.3:The normalised wavelet power spectra of yearly mean ET (a) and Precipitation (b) variability during 2001 to 2018 at Keiskammahoek Irrigation Scheme 64

Figure 4.4: Wavelet coherence between ET and Precipitation for period (2001 to 2018) with phase relationship depicted by the arrows. 65

Figure 4.5: Long-term trend of monthly SF, Sandile Dam MV, ET (a) and (P), MV (c), SF (d). 69

Figure 4.6: Sequential Mann-Kendall (SQ-MK) test for ET and P data for the period from 2001 to 2018. 72

Figure 4.7: Model flow chart 75

Figure 4.8: Auto-Correlation Function (ACF) of residual as an ideal fitted model for data series of ET from 2001 to 2018. 76

Figure 4.9: Histogram of residuals as a best fitted model for data series of ET from 2001 to 2018. 76

Figure 4.10; ARIMA and ANN training using data from (2001-2018) 77

Figure 4.11: Predicted evapo-transpiration for 3-year period from 2016 to 2018 using the ARIMA model. 78

Figure 4.12: Scatter plot between observed and predicted values using ARIMA model (with validation period of 2016 – 2018). 79

Figure 4.13: The actual data from 2001 to 2015 and the predicted evapotranspiration results using NNAR models from 2015 to 2018. 80

Figure 4.14: Scatter plot between observed and predicted ET using NNAR model (with validation period of 2016 to 2018).. 80

Figure 4.15: Observed ET versus predicted ET from 2001 to 2015 using the Hybrid (ARIMA-NNAR) model (black line indicates observed data and blue line indicates predicted data). 82

Figure 4.16: Scatter plot between observed and predicted ET using Hybrid (ARIMA-ANN's) model (with validation period of 2016 to 2018)... 82

Figure 4.17: Shows predicted ET data from 2016 to 2018. The black line indicates the prediction using the averaged models (ARIMA, NNAR and Hybrid) models 84

Figure 4.18: Scatter plot depicting observed ET and the predicted ET using the averaged models. 84

Figure 4.19: Observed ET from 2001 to 2018 and predicted ET for five years (2018 to 2023) using the Hybrid (ARIMA-ANN) model. 87

LIST OF TABLES

Table 4.1: Pearson correlation coefficient for NDDI, ET, NDVI, NDWI, SF and Precipitation (P). 66

Table 4.2: Long-term trend of monthly SF, Sandile Dam MV, ET (a) and (P), MV (c), SF (d). 70

Table 4.3: Multi-Linear Regression (MLR) model where ET is the dependent variable and SF, P (P), NDVI and NDWI are independent variables. 73

Table 4.4: Comparison of ARIMA and ANN: RMSE, MAE, MAPE and R 78

Table 4.5: Comparison of the ARIMA, ANN, Hybrid and Combined Models: RMSE, MAE, MPE, and R. 85

ACRONYMS AND ABBREVIATIONS

ACF	Auto-Correlation Function
ADM	Amathole District Municipality
AIC	Akaike's Information Criteria
ALM	Amahlathi Local Municipality
ANN	Artificial Neural Networks
ARIMA	Autoregressive Integrated Moving Average
BCMM	Buffalo City Metropolitan Municipality
BFAST	Break for Additive Seasonal and Trend
CHJ	Corrected Fanses Haise
COI	Cone on Influence
DWA	Department of Water Affairs
EC	Eastern Cape
ESA	European Space Agency
ET	Evapo-transpiration
FIS	Fuzzy Inference System
FPM	FAO Penman Monteith
FBC	FAO Blaney Criddle
FRM	FAO Radiation Making
GRNN	Generalised Regression Neural Networks
HDUG	Hydrological Date User Group
HRG	Hargreaves
HS	Hargreaves Samani
LDPE	Low-Density Polyethylene
LM	Lavenberg Marquardt
MAE	Mean Square Error
MAFPE	Mean Absolute Forecasting Percentage Error
MAPE	Mean Absolute Percentage Error
MERRA-2	Morden Era Retrospective
MK	Mann-Kendall
MLP	Multi-Layer Perception
MLR	Multi Linear Regression

MSE	Mean Square Error
MV	Monthly Volume
NASA	National Aeronautics and Space Administration
NDDI	Normalised Difference Drought Index
NDVI	Normalised Difference Vegetation Index
NDWI	Normalised Difference Water Index
NNAR	Neural Network Auto-Regression
P	Precipitation
PACF	Partial Auto-Correlation Function
PT	Priestley Taylor
R	Correlation Coefficient
RBFNN	Radial Basis Function Neural Network
RMSE	Root Mean Square Error
SA	South Africa
SD	Sandile Dam
SF	Stream Flow
SQ-MK	Sequential Mann-Kendal
TSE	Theil-Sen Estimator
TW	Thornthwaite
USA	United State of America
WSEE	Weighted Standard Error of Estimate

CHAPTER 1: INTRODUCTION

1.1 INTRODUCTION

Climate change is a global phenomenon which causes unusual variations in the atmosphere with the resultant effect of changes in temperature and rainfall patterns, changes in normal wind direction patterns, and increased intensity and frequency of extreme events, such as droughts, floods, and cyclones (kahsay & Hansen, 2016) Click or tap here to enter text.. Climate change, as defined by IPCC (2007), is “the alteration of climate state where changes in the mean and variation of its properties can be easily identified” and it can manifest itself in three scenarios, namely, mild, moderate and severe according to the uncertainties associated with it (Connor, Schwabe, King, & knapp, 2012). It is a global phenomenon that is expected to affect agricultural water availability with resultant effects, such as reduced crop production, increased food prices, and food insecurity (Calzadilla, Zhu, Rehdanz, Tol, & Ringler, 2014).

According to Ziervogel, et al., (2014), the increase in annual temperatures in South Africa by at least 1,5 times of the average 0.65 degrees has led to climate being a key concern. They further suggest it as posing a significant treat to South Africa water resources, food security health, infrastructure, as well as ecosystem services and biodiversity. The growing impact of cli-mate change has major implications for South Africa Healthwise especially for more vulnerable groups even though there are policies promoting an ambitious renewable energy program, South Africa’s response to climate charge is hardly hampered by uncertainty in policies and corruption (Chersich, Wright, Venter, Rees, & Erasmus, 2018). In Eastern Cape province of South Africa livestock farming in an important agricultural practice and it considered as the wealth of famers despite their education status (Mandleni & Anim, 2011). According to the conclusion of (Todaro & Snith, 2012), livestock farmers suffer a greater impact from climate change. South Africa suffers from scarcity of water as the demand for water resources increases with the increase in population. If the country wants to sustain economic development, urgent needs must be in place to protect the quality of the resources whilst fighting to meet the problem of water scarcity (Chersich, Wright, Venter, Rees,

& Erasmus, 2018). Most of the land in Eastern Cape is used for agriculture with around 35% of households involved in agricultural activities, however the extreme drought conditions over decades have negative im-pact on these famers (Nkondo, zyl, Keuris, & Shrener, 2012). In South Africa, irrigation accounts for over 55% of the total available consumptive freshwater (Mishra & Singh, 2011). South Africa falls within the semi-arid region, where the evaporation rate is more than the precipitation rate (Calzadilla, Zhu, Rehdanz, Tol, & Ringler, 2014).

Evapo-transpiration (ET) is a combination of evaporation and plant transpiration, and this process happens from the earth's surface to the atmosphere. Evaporation accounts for the conveyance of water from the soil, canopy intervention and water bodies, whilst transpiration takes place during water loss from plants as they lose water through stomata in their leaves (eurk, 2012). ET is one of the crucial elements of the hydrological cycle, which expedites constant precipitation, through the process of condensation, back to the land and ocean surface as rainfall. It is also crucial for the conveyance of minerals and nutrients necessary for plant growth. ET creates a favourable cooling method for plant canopies in many climates and has direct association with Latent Heat Flux Effect (LE) on the earth's energy and water balance (Ramoelo, et al., 2014). ET is essential in irrigation water planning and management. Therefore, evapo-transpiration is significant as one of the main constraints on irrigation development in developing countries, as well as in semi-arid regions of the world (Traore, Wang, & Kerh, 2008)

Accurate prediction of evapo-transpiration is essential for irrigated agriculture as it informs proper planning and contributes positively to the daily management of the irrigation scheme. Moreover, determining the perfect timing and amount of water needed for irrigation is important for effective management of water absorbed by crops (Kishore & Pushpalatha, 2017). In a study by Bachour, (2013) it was reportd that agricultural managers have been making timely predictions of evapo-transpiration for effective use of water, as it is the pre-condition for effective water management.

However, it is not easy to measure ET in arid and semi-arid regions because of the small magnitude of ET flux (Ramoelo, et al., 2014). Because of the difficulty in getting all the relevant data to use the widely known Penman-Monteith approach, alternative reliable and powerful prediction approaches are used to examine the non-linear trends related to the prediction variables for the rate of evapo-transpiration (Ghorbani, Kazempour, Chau, Shamshirband, & Ghazvinei, 2018)

1.2 PROBLEM STATEMENT

There is a shortage of water in various parts of the world because of an increase in water demand due to growth in population, energy use, industrial sectors, growth in agriculture and increase in climate change (Mishra & Singh, 2011). South Africa is also classified as a water-stressed country. The current, major problem at the Keiskammahoek Irrigation Scheme is not unrelated to the problem of global climate change. The farm involved in the Keiskammahoek Irrigation Scheme has 750 hectares of land with 600 irrigated sections. The significant decrease in the levels at the supply dam on top leaves the farm with no other choice but to close other irrigated sections of the farm and this severely affect the dairy production, which is a huge problem considering the number of livestock the farm owns. In the absence of accurate information on ET estimation, as well as a lack of prediction of the potential evapo-transpiration, which is the main indicator of irrigation water requirement, it will be difficult to plan for an efficient water management system on the irrigation scheme. A study by Santos, Lorite, Tasumi, Allen, & Fereres, (2007), in GenilCabra Irrigation Scheme of Spain proved that estimated ET does provide accurate irrigation scheduling guidelines where the irrigation water is often limited. Their study also assisted in identifying certain agricultural fields which had problems in water management. Proper estimation of ET assists in getting the actual water that has to be used by the plants, as the plants only uses water retained in root zone, whilst other water below the root zone is lost (Hsiao, Steduto, & Fereres, 2007). Therefore, Estimation of ET will assist in determining the exact water that has to be used to irrigate the site in order to overcome the problem of limited water around the keiskammahoek irrigation Scheme.

1.3 AIM

The main aim of this study was to predict evapo-transpiration (ET) for optimum water use in agriculture at the Keiskammahoek Irrigation Scheme, Eastern Cape, South Africa.

1.4 OBJECTIVES

In this study, the above aim was achieved through the following objectives:

a) To collect and analyse the time series such as ET, Precipitation, Normalised Difference Vegetation Index (NDVI), Normalised Difference Water Index (NDWI) and Normalised Difference Drought Index (NDDI).

b) To develop Autoregressive Integrated Moving Average (ARIMA), Artificial Neural Network (ANN) and Hybrid (ARIMA-ANN) models for time series modelling using R-project language and predict eva-potranspiration using these models

c) To analyse and compare the performance of all the models.

1.5 SIGNIFICANCE OF THE STUDY

In the past, historical data have been used to predict ET in different regions all over the world. This study is significant because it not only predicts evapo-transpiration at the Keiskammahoek Irrigation Scheme, but also introduces a novel model that can be used to predict evapo-transpiration for general optimum water use in Agriculture. Currently the Scheme uses irrigation for their crop and there is not estimate of how much water is needed and proper prediction of ET will assist in overcoming the over-use of limited water. The predicted evapo-transpiration results are relevant and useful for irrigation/water resource managers. Hence, this farm will be able to plan effectively and implement their irrigation scheduling and water usage efficiently. With the scarcity of water caused by climate change and a decrease in precipitation rate in the Eastern Cape, it is advisable to schedule and manage water for irrigation to promote water conservation.

1.6 STRUCTURE OF THE DISSERTATION

This dissertation is divided into five chapters as follows:

- Chapter 1: contains an overview of the study, and an introduction of the aims and objectives of the study, the problem statements, the study area and the significance of the study. The chapter concludes with the envisioned outcomes of the study.
- Chapter 2: contains an in-depth review of the past and current literature on evapo-transpiration and other contributing factors relevant to the topic of the study.
- Chapter 3: contains an explanation of the research methodology used to conduct this study. The data collection procedures, prediction tools and the mathematical models used in the study are discussed in this chapter.
- Chapter 4: in this chapter, the analysis of the results of this study is presented, including a thorough discussion of the results.
- Chapter 5: this chapter contains the summary and conclusions of the study based on the results obtained. Recommendations for future studies are also provided in this chapter.

CHAPTER 2: LITERATURE REVIEW

2.1 STATE OF IRRIGATION IN SOUTH AFRICA

The introduction of irrigation in South Africa began after the arrival of European settlers. However, many researchers have fully explained its evolution, emphasising the initial inequality that existed amongst irrigation policies between black- and white-owned irrigation schemes. The publication of the Tomlison Commission Report in 1955 led to the development of many irrigation schemes and recommendations made in the report had a major impact on land-use patterns and development of irrigation in black-inhabited rural areas as well as settlements (Perret, 2002). Approximately 30% of crop production in South Africa is under irrigation and this makes irrigation the leading user of water. South Africa is a water-stressed country. Although there has been a decrease from 80% to 50% in the use of water for irrigation, there is still much need to encourage effectiveness of water use in irrigation (Fanadzo, Chiduza, Mnkeni, Stoep, & Stevens, 2010). Irrigation is practised on an approximately 1.5 million ha and an estimated 0.26 x10^6 ha are affected by salinisation. Most of the irrigated areas are in very dry areas that are unsuitable for rain-fed farming. Irrigated agriculture is possible in these areas, although it is the main consumer of water (Annandale, Stirzaker, Singels, Laan, & Laker, 2011). According to Perret, (2002). the irrigation of land and watering of livestock consume a total amount of 52% of the total annual use of water in South Africa.

ET is essential for irrigation water planning and management. Therefore, ET is significant as one of the major constraints on irrigation development in semi-arid regions of the world (TRAORE, WANG, & KERH, 2008). The difficulty in obtaining all the relevant data to use the widely-known Penman-Monteith approach means that alternative reliable approaches to prediction are used to examine the non-linear trends related to the prediction of variables for the evapo-transpiration (Ghorbani, Kazempour, Chau, Shamshirband, & Ghazvinei, 2018). Therefore, in this study it is expected that these alternative models will be able to provide improved estimation of

ET, which will assist water managers to have the effective irrigation scheduling systems.

2.2 HYDROLOGICAL CYCLE

The hydrological cycle traces the greatest movement of any kind of substance on Earth. The hydrological cycle instigates climate in many ways through moisture and heat exchange between the earth's surface and atmosphere, which impacts the thermodynamics as well as dynamics of the climate system. Through the vapour process, liquid, ice, water, and snow are affected during transition phases in which water is a key opposing element in cooling as well as heating the environment (Chahine, 1992). Various activities by human beings release small particles into the atmosphere and these human aerosols advance the scattering and ingestion of solar radiation. This leads to brighter clouds with less ability to release precipitation, which leads to an extreme decrease in the quantity of solar irradiation arriving on the earth's surface, a corresponding multiple in atmospheric solar heating, modification of temperature structure in the atmosphere, freezing of rainfall and reduced disposal of pollutants, which leads to a weaker hydrological cycle (Ramanathan, Crutzen, Kiehl, & Rosenfeld, 2001).

Figure 2.1 shows the major reservoirs of water and fluxes where the ocean dominates the reservoir globally with 97% of the water worldwide, the atmosphere holding 0.001 % and the remainder being locked in ice caps, which are the snow, and storage underground. Chahine, (1992), further described the hydrological cycle as being global, hence, both oceans and continents exchange water, and stated further that, over the oceans, evaporation exceeds precipitation, and this variance has an impact over the land.

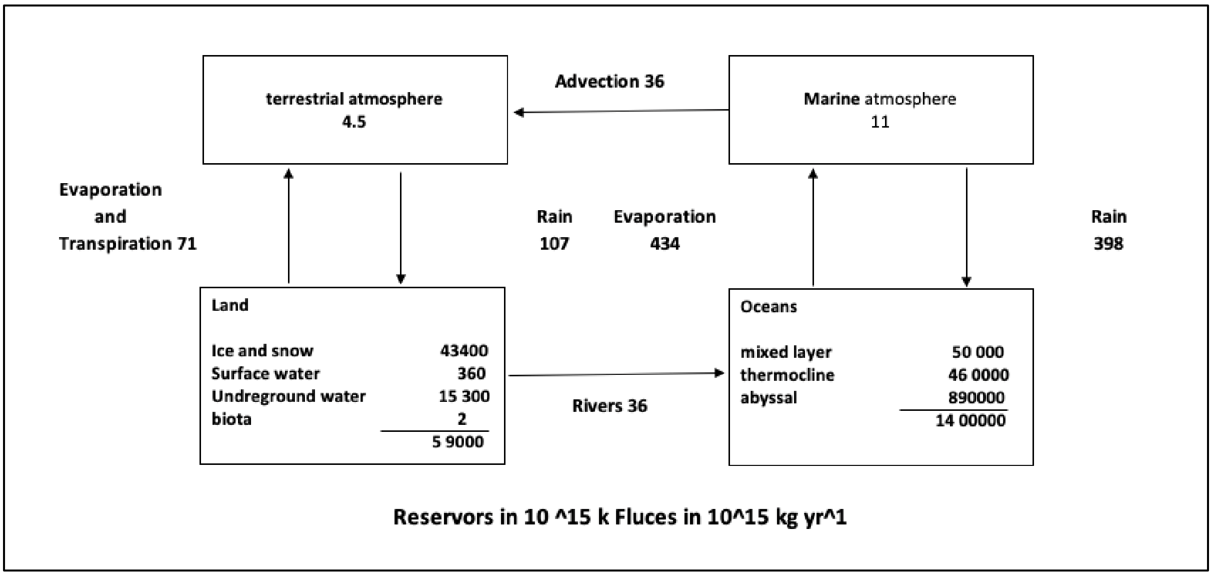


Figure 2.1: Main reservoirs and fluxes of water (*Source: Chahine, 1992)*

This subject has been revisited recently by Koutsoyiannis, (2020) who was motivated by the availability of various hydrological data and advanced technology. Koutsoyiannis, (2020) processed information from gridded ground information, re-analyses of available data and satellite data and found that widely created hypotheses could not be confirmed. Koutsoyiannis, (2020) established further that water on sea and land had decreased significantly compared with the general figures presented in literature with extreme inconsistency on climate timescale, in agreement with the stochastic dynamics of Hurst-Kolmogorov. Koutsoyiannis, (2020) found a clear anthropogenic signal within the hydrological cycle, which unfolded to be the maximum exploitation of groundwater affecting the sea gauge rise.

2.3 EVAPO-TRANSPIRATION

Evaporation is transportation of water vapour from the soil, canopy interception and water bodies into the atmosphere, whilst transpiration takes place from plants to the atmosphere through stomata in their leaves (Khoshhal & Mokarram, 2012). Allen, Pereira, Raes, & Smith, (1998) described ET as the loss of water through a combination of evaporation and transpiration. ET is considered to be one of the most important elements of the hydrological cycle. It is also crucial for the dispersion of minerals and nutrients necessary for plant growth. ET creates a favourable cooling process for plant canopies in many climate conditions and has direct implications for

Latent Heat Flux Effect (LE) on the earth's energy and water balance (Ramoelo, et al., 2014). Jovanovic, Mu, Bugan, & Zhao, (20215) described ET as being the key procedure within the hydrological cycle and one of the most difficult variables to determine, particularly in arid and semi-arid regions where a huge percentage of the little rainfall is returned to the atmosphere. Semi-arid regions are areas where vegetation is always under water stress and various plants adapt themselves in different ways to survive in the constant drought situation. Jovanovic, Mu, Bugan, & Zhao, (20215) added that ET is predicted to exceed 60% of the total rainfall world wide, with the potential to reach 100% in arid regions.

ET is one of the meteorological parameters that is vital in the hydrological cycle. These parameters are part of the regional climate, and they are always the result of differences caused by factors such as rain, water, wind etc. and these parameters are determinants of drought (Asadi, Vahdat, & Sarraf, 2013). In irrigated agriculture, it is important to quantify ET accurately in order to have an accurate estimation of the crop water requirement which is necessary for irrigation scheduling. This assists in formulating options for management of irrigation water to maximise production, depending on the type of crop, type of cultivar, and type of soil. Therefore, it is essential to predict ET effectively in irrigated agriculture in order to attain a comprehensive picture of the water cycle (Dutta, et al., 2016) in order to manage scarce resources for irrigating crops (Anapalli, Fisher, Reddy, Rajan, & Pinnamaneni, 2019).

2.4 EVAPORATION

The process of evaporation happens when liquid water is converted into water vapour and transferred in this form into the atmosphere. It is a naturally occurring process with energy input from the sun or atmosphere, determined as the rate of energy of vapour defusing away from the surface. Molecular diffusion is one of the physical processes, with turbulent diffusion that occurs at subsequent stages during the path of this transference. These processes are naturally ground-based, and they can be expressed in physical terms by models that define the impact of molecular and turbulent diffusion resistances to the dissolution of energy from the atmosphere or the sun (Shuttleworth, 1979). Spontaneous evaporation is stimulated by the heat supplied from the atmosphere and by circulation of water within the plant blocking the decision of leaf tissue. The process is similar to commercial transformation in which a wet surface releases water vapour into its environment in response to heat. The evaporation from plants is caused by heat supplied by solar radiation, from soil conduction and through turbulence conveyed from the atmosphere (Montheith, Evaporation and environment, 1965). According to Wilfried, (2005), evaporation can be explained as “the phenomenon by which a substance is converted from the liquid or solid state into vapor and in case of a solid substance, the phenomenon is often referred to as sublimation”. Wilfried, (2005) described evaporation further as a phase in the hydrological cycle that is key, and the cycle is completed when water vaporises into the atmosphere.

The physicality of natural evaporation is a key variable of oceanography, hydrology, and meteorology (Montheith, Evaporation and surface temperiture, 1981).

> Evaporation of water plays an important part in most human activity designed to satisfy basic needs. While, in the technology that produces clothing and shelter, evaporation processes are under some measure of control, in the open-air operations that lead to production of food and drink, evaporation is usually beyond control and amounts and rates are often very difficult to estimate. Hence our problem. Before passing on to the "food" aspect, we might note that, in the "drink" aspect, the water engineer regards evaporation of water as a loss, whether it occurs from reservoirs, from streams, from bare soil or from land carrying crops. Water supply and agriculture must often be in conflict. The farmer's attitude varies from time to time. While waiting to start spring cultivations, he will regard evaporation as an essential preliminary to seed-bed preparation; later, the evaporation from bare soil between the growing plants will tend to be regarded as a waste of water that would be more profitably used in passing through the plant. At all times he will regard transpiration as helpful, because it is an essential condition of plant growth (Penman, 1956).

2.5 TRANSPIRATION

Transpiration is the movement of water from soil through a vascular plant and into the air, which happens by a passive and quick-drying mechanism explained by the Cohesion Tension Theory. In the theory it is stated that transpiration is the decreasing of water by evaporation that reduces the pressure of liquid water inside the leaf in relation to atmospheric pressure and such decreased pressure withdraws liquid water from the soil up the plant stem for the purpose of maintaining hydration (Wheeler & Stroock, 2008). Schlesinger & Jasechko, (2014), described transpiration as the main factor of rainfall and micro-climate. It is the response to increasing atmospheric carbon dioxide and it accounts for 60% to 80 % of the land evapo-transpiration.

Jasechko, et al., (2013) discovered evidence that, locally there can be open-water evaporation occurring on a greater scale than transpiration. However, the fraction of the sum of evapo-transpiration shown by evaporation is extremely limited in lesser areas of open water on continents on Earth, which is approximately 3% globally. Consecutively, Jasechko, et al., (2013) discovered transpiration to be responsible for greater than two-thirds of the total evapo-transpiration of surface water as well as being responsible for large amounts of evapo-transpiration in desert catchments

2.6 IRRIGATION SCHEDULING

Irrigation scheduling can easily be justified as strategic programmes of irrigation use, indicating date and quantity of water, to attain a goal of maximum production. Efficient irrigation scheduling can lead to higher profits, preventing waste of water, preventing plant stress, increased water utilisation productivity and achieving higher earnings (Singels & Smith, 2006). Singels & Smith, (2006), suggested further that these objectives are attained mostly by trying to keep the water content in the soil within a preferred range, given all the constraints of irrigation systems and soil. This process of irrigation scheduling requires accurate knowledge of the daily crop water requirement which, in turn, is dependent on the knowledge of daily ET. Olivier & singels, (2015), also described in their study that accurate irrigation scheduling is one of the key factors that could improve effective use of water through saving or enhancing harvests. Many researchers suggest that there are several scheduling instruments available, ranging from comparatively uncomplicated methods to quantify soil content precisely to modern crop simulations. Irrigation scheduling, consequently, performs a major function in influencing crop production from the use of water, and it is considered to be the performance gauge used to define the correlation between water employed and agricultural productivity (Annandale, Stirzaker, Singels, Laan, & Laker, 2011).

ET, a combination of evaporation and transpiration, is a process of conveying the moisture from land to the atmosphere, directly from the soil and through plants. It is, therefore, very important to predict this value in order to quantify the crop water requirement and the water available in the soil. In this way, it will assist farmers to manage water that is required by crops efficiently (Caminha, da Silva, & da Rocha, 2017). Evapo-transpiration happens from both rainwater and irrigation water. Thus, it is necessary to maintain sufficient soil moisture in the soil depending on the type of soil and type of crop. Irrigation scheduling, which is the determination of “when” and “how much”, becomes critical in irrigated agriculture, where accurate ET estimation will give an assurance of “how much” water is required in an irrigation scheme, design, and project planning (Kishore & Pushpalatha, 2017).

2.7 STATE OF DROUGHT IN EASTERN CAPE

Botai, et al., (2020) described drought conditions as extremely intense and ubiquitous, adverse weather phenomena, which manifest as an enlarged period of deficiency in precipitation conditions compared with normal conditions. Botai, et al., 2020) described the drought manifestation through decreased precipitation as a meteorological drought, which results from imbalance of water availability, reflected in inconsistent, average precipitation. Drought is one of the common natural disasters that has affected the farming community on a large scale in South Africa. The results of this disaster and increased floods are a loss in economy, and the number of people seriously affected, based on mortalities indicated by the data on a chart from the Centre for Research on Epidemiology of Disaster (2011), show that it has a major impact. Many communities are vulnerable to drought with an estimated 60% of sub-Saharan Africa (SSA) and South Africa being part of the affected countries, with an estimated 65% of the region getting less than 500mm of rainfall in a year.

According to Ngaka, (2012), the current drought signs indicated in the Eastern Cape are not new and many researchers have indicated this going back to 1992. Jury & levey, (1993) described the Eastern Cape as a province that has suffered drought with a negative effect on agricultural production and impact on water resources of the province. According to the analysis by (Jury & levey, 1993), drought occurred at an average of 2.45 times in 18.2 years, specifically during March and that affected dam levels which were at 30% capacity in early 1992. Jury & levey, (1993) further reported this excessive drought as being the basic influence on declined economy because of its effect on agricultural production and scarcity of water. Botai, et al., (2020) investigated drought characteristics based on joint distribution of severity and duration. Botai, et al., (2020) used time series data from a 6-month and 12-month, Standardised Precipitation Index, estimated from monthly rainfall over the previous five decades. The output of the study by (Botai, et al., 2020) showed a dependency structure indicating that the study was experiencing decreased probability of drought and increased probability of drought severity. Graw, et al., (2017) distinguished four drought types as being agricultural drought, hydrological drought, socio-economic drought and meteorological drought, and identified agricultural drought as having an influence on a decrease in water for plant growth, and meteorological drought, characterised by the Standard Precipitation Index (SPI), as having an influence on decreased precipitation.

The experience of the Eastern Cape is comparable with that of the rest of the world which suffers from the impact of drought on agriculture, resulting in increased food prices and a drop in economy. According to (Graw, et al., 2017), the climate of this province ranges from mild warm to sub-tropical temperature conditions which exposes the province to droughts that can occur throughout the year, making agricultural crops vulnerable to droughts in summer. These variations in drought affecting the southern part of Africa are associated mostly with the *EL Nino* phenomenon that leads to reduced average rainfalls. The drought conditions during 2015/2016 were triggered by *EL Nino* and they affected mostly agricultural land in the Eastern Cape, where an estimated 35% of households are engaged in agricultural activities, with crop farming forming 32% and 65% being involved in agriculture. These large proportions indicate the importance of grassland in this province, which is affected by overstocking, which is an even greater problem for South Africa. Graw, et al., (2017), further characterised communal land, where traditional management is practised, depending on rainfall, and commercial land where large-scale farming is practised, using innovative strategies, including irrigation activities. Abdel-Hamid, Dubovyk, Graw, & Greve, (2020) investigated the effects of droughts that affect communal and commercial grasslands in the Eastern Cape. Based on the study, the Eastern Cape was reported to be a province that is highly subject to climatic variations, which is evident in the variations in vegetation. Abdel-Hamid, Dubovyk, Graw, & Greve, (2020, reported further that communal grassland areas feel impact of drought more than commercial land because of management activities, such as irrigation, which assist in decreasing the impact of drought whilst subsequently improving the durability and output of the ecosystem.

The consequences of periodic drought in this province have significant implications for the agricultural sector. This severe drought since 2015, with certain recoveries in 2020, had negative effects on various sectors, including livestock, and this drought state has been reported in the media as being unprecedented (Archer, et al., 2022). The results reported by scholars indicate that the ongoing concern of occasional drought is not necessarily unprecedented, but very intense with highly critical, local effects. Scholars reported further that the loss of production and income are likely to continue to affect the agricultural sector in the Eastern Cape.

Various studies have been conducted in the Keiskammahoek area regarding the drought and its impact on agriculture. Haindongo, (2009) investigated the influence of vegetation stress in part of the Keiskamma catchment and found that vegetation in this region is highly stressed based on the two study techniques used: discrepancy in soil moisture, and conditions on the soil surface have severe impacts on vegetation stress. Based on the study, it was concluded that the existence of stress in vegetation would continue in the Keiskamma catchment. Evapo-transpiration is one of the key variables which have been used to assess drought globally.

Vicente-Serrano, Van der Schrier, Begueria, Azorin-Molina, & Lopez-Moreno, (2015) assessed the contribution of evapo-transpiration and Precipitation, referred to as (ET_0) and (P) respectively, to the well-known drought indices which include: the Palmer Drought Severity Index (PDSI), Reconnaissance Drought Index (RDI), Standard

Precipitation Evapo-transpiration Index (SPEI) and Standard Palmer Drought Index (SPDI). The results indicated a high influence of (ET_0) and (P) on the drought indices, including a sensitivity series with regard to change and variance, and further suggested the use of these two indicators to assess the climate change effects on drought. ET has also been assessed as a drought index.

In a study by Haindongo, (2009), actual evapo-transpiration was used, with estimated ET and predicted ET from the Boucher hypothesis and the structure of the Standard Precipitation-Evapo-transpiration Index to establish a drought index based fully on ET. This investigation gave similar results for assessing drought as the results of the Palmer Drought Severity Index (PDSI) and Standard Precipitation Index (SPI). This indicates a greater importance of ET in assessing drought, and its prediction at Keiskammahoek will also assist other researchers to determine possible signs of drought.

2.8 DROUGHT ASSESSMENT INDICES

2.8.1 NORMALISED DIFFERENCE WATER INDEX

Normalized Difference Water Index (NDWI) is one of the most widely used remote sensing parameters. This method, which absorbs water at 1240nm, was developed by (Gao, 1996), in order to quantify the water content of vegetation using remote sensing reflectance. The index has been used by various researchers to predict the water content of vegetation (Chapungu & Nhamo, 2016). "NDWI is a measure of liquid water molecules in vegetation canopies that interacted with the incoming solar radiation. It is less sensitive to atmospheric scattering effects than NDVI" (Gao, 1996). NDWI is the ratio between dual separate bands, which improves the water spectral signal by distinguishing the reflectivity among several wavelengths and eliminating enormous sections of noise elements in diverse wavelength areas (Campos, Sillero, & Brito, 2012).

It has been used successfully to define surface water characteristics (Zhang & Wylie, 2009), in many regions, the model has been used successfully on agricultural crops and it has proven to be the best indicator for the prediction of water content in corn

(Jovanovic, Garcia, Bugan, Teich, & Rodriguez, 2014). Therefore, it is important to assess this variable in Keiskammahoek as the availability of water has an impact on evapo-transpiration occurring at the study site.

2.8.2 NORMALISED DEFERENCE VEGETATION INDEX

This index is a widely adopted technique to examine the remote sensing reflectance and it assesses the vegetation greenness or health. The higher the value of this index indicates that there is a higher level of photosynthesis activity in the particular study area, while the lower the value of this index reflects the moisture stress in vegetation, which might be caused by low precipitation (Kumar, Tanwar, & Singh, 2017). The index best explains the vegetation activity and energy levels and is used to predict spatial variables of moisture and surface and cloud temperatures (Jovanovic, Garcia, Bugan, Teich, & Rodriguez, 2014) .

In a study by Cihlar, laurent, & Dyer, (1991) it was proved that there was a significant correlation between ET and NDVI. Irrigation in Keiskammahoek is under water stress and the NDVI was measured in order to assess the greenness of crops over the study period in relation to ET over the same 18-year study period.

2.8.3 NORMALISED DIFFERENCE DROUGHT INDEX

The Normalised Difference Drought index (NDDI) is one of the most used indices to monitor and map the number of droughts that are associated with environment and climatological conditions and is one of the indices extracted from satellite images. An index is an appropriate tool for recognising the effect of drought conditions and water availability on the health of vegetation in semi-arid areas, hence, it includes NDVI and NDWI (Orimoloye, et al., 2019). According to Gulacsi & Kovacs, (2018), the index can be used as a sensitive drought monitoring technique for agriculture, water management, and conservation of natural areas. Different satellite-based indices, such as NDWI and NDVI, have been used in the past to assess drought, but these indices could not give the appropriate water content in vegetation, thus the normalised difference drought index (NDDI), which is a combination of NDWI and NDVI, was developed as the drought index (Lee, et al., 2016). Lee, et al., (2016) argued further that greater values of NDDI indicate drought conditions and smaller values of NDDI indicate non-drought conditions.

The Keiskammahoek Irrigation Scheme is situated in a water-stressed environment, assessing variables, such as NDDI, to assist in checking whether there are any of drought. According to Lee, et al., (2016) evapo-transpiration increases when there is low precipitation and higher temperatures because of drought, which reduces moisture in the soil, thus triggering vegetation stress.

2.9 EVAPO-TRANSPIRATION PREDICTION TECHNIQUES

Different researchers have predicted evapo-transpiration (ET) in various regions Yang, et al., (2006) used a machine learning technique, based on a support vector machine (SVM), to predict ET using data collected from 2000 to 2004. The prediction model required two sets of data: ground-measured ET for machine training and other data for application in the model. The results showed that machine learning techniques with data collected from the ground and remotely sensed inputs can predict ET over the coterminous in the United States. However, the model had a limitation in that (Yang, et al., 2006); proposed the use of bigger ET ground-observation data, as the knowledge learned was determined to be not easily understandable to humans. However, (Yang, et al., 2006) found the final, improved SVM techniques to be accurate enough to predict ET with ground data collected. Based on the accuracy of the machine learning technique applied, (Yang, et al., 2006) concluded that SVM-based prediction of ET is important, and the model could be used for the purpose of documenting the hydrological models on different scales.

In a study by Patel & Balve, (2016), Fuzzy Logic System was used to predict ET at Takali Meteorological Station in the Nashik District of Maharashtra State located in India. They used daily climate parameters from January to June 2013, obtained from the Hydrological Date User Group (HDUG). To determine the ET values, the Penman-Monteith technique was used to analyse the daily data collected. Using the MATLAB Fuzzy Model, the Fuzzy Inference System (FIS) was developed and used to predict ET, and the results were compared with the FAO-56 Penman-Monteith technique. The results proved that the Fuzzy Inference System has the capability to estimate ET, given enough daily data to satisfy the technique.

Pandorfi, et al., (2016) employed Artificial Neural Networks to predict evapo-transpiration of greenhouse-grown sweet paper vegetable at the Department of Biosystems Engineering of Luiz de Queiroz College of Agriculture of the University of Säo Paulo, in the city of Piracicaba, in Brazil. Data from 135 days during September 2013 to February 2014 were used, constituting temperature and relative air humidity, solar radiation and wind speed, and ET was determined using the Lysimetric technique. The data were divided into three sets for training, testing, and validation. The results obtained from the use of ANN permitted the development of ET patterns in the green-house environments even though the performance of the specialist model system had to be checked regularly. The results proved that ANNs were the best for predicting ET and the model can be used for decision-making.

2.10 TIME-SERIES PREDICTION

According to Anderson, (1995) "A time series of length n is an ordered sequence of n observations recorded at equi-spaced instants and denoted by z_1, z_2 ……,z_n,". A time-series is several observations that happened consecutively through time. These observations might be continuous in time or seen as a separate set of time points. The two types of time series, by convention, are called continuous or discrete time series. Time series have four objectives: description of data in a summary way; finding the most suitable way to analyse a data-generating process, called Modelling; time series used to predict the future variables of the series, called Predicting; and, lastly, time series to control, which controls the action of an inserted process (Chatfield, 2000). Time series is a crucial component of predicting, in which previous measurements of a similar variable are obtained and examined to generate a model by labelling underlying correlations (Zhang G. P., 2003). There are linear and non-linear types of time series analysis. Tsay, (2005) explained that linear time series models give a neural outline to learn the structure of the dynamics of a certain series. A stationary time series has a mean and the auto-correlation must be constant over some time (Gautam & Sinha, 2016). These models typically need a huge set of data collected endlessly over numerous sequences of time (Dingwell, 2000). In the current study time-series prediction was employed to predict evapo-transpiration at the Keiskammahoek Irrigation Scheme. According to (Hamdi, Bdour, & Tarawneh, 2008), use of time series prediction has become common recently because of the complexity involved in direct measurement of evapo-transpiration and the cost involved in the process. Alternative time-series models using historic data can be the solution for predicting ET (Hamdi, Bdour, & Tarawneh, 2008). These models are very accurate, and their dominance has led to them being used regularly in prediction (Zhang G. P., 2003).

2.10.1 Auto-regressive Integrated Moving Average (ARIMA) Models

ARIMA model is one of the most widely used models because of its statistical properties and it can be used in different ways such as pure auto-regressive (AR), pure moving average, and combined ARIMA series (Kishore & Pushpalatha, 2017) .This Box-Jenkins modelling approach is one of the most used time series because of its flexibility, even though it cannot predict non-linear relationships because its linear correlation structure is presumed among the time values (Zhang, Zhang, & Li, 2016).

ARIMA models have been used in the past and they were able to predict accurately. In a study by Gautam & Sinha, (2016), seasonal auto-regressive integrated moving average was used to model and predict evapo-transpiration to assist water managers in decision-making and managing water in Bokaro District, Jharkhand, India. Gautam & Sinha, (2016) stated that, in time series, by “studying the past, better decisions for the future can be made”. The models and prediction results of (Gautam & Sinha, 2016), were crucial in developing local and national policies and for irrigation scheduling. Even though data over 24 months were used to model and predict in the study to obtain valid results, (Gautam & Sinha, 2016), argued that the prediction technique used would allow for more months even though the accuracy might be decreased.

In the study by Valipour, Banihabib, & Reza Behhahani, (2013), monthly release during 42 years of flow from reservoirs at Taleh Zang station was used and ARMA, ARIMA, and Autoregressive ANN models were used to predict for 5 years. ARIMA models produced much improved results in both training and predicting because they take fixed time series into consideration. ARIMA was also used by Valipour, Banihabib, & Reza Behhahani, (2013), to examine and equate the performance of ARIMA and ANN, and Penman-Monteith equations, and ARIMA and ANN were found to be less accurate than the method based on weather prediction. ARIMA was one of the models used to develop the univariate model to predict the "summer monsoon (June-August) rainfall over India". These models were adopted after realising the causality and non-stationary time series in the data used from 1871 - 1999 and the trend and randomness within the time series were investigated. Even though ARIMA was not eligible because of the large value of Willmott's Index, it was suggested as the alternative model (Chattopadhyay & Chattopadhyay, 2010).

In a study by Contreras, Espinola, Nogales, & Conejo, (2003), ARIMA models were used to predict the next-day's electricity prices and they were able to present their ARIMA results to Spain and California. Feng, et al., (2016), used ARIMA successfully to model monthly and annual rainfall, reference evapo-transpiration, and Dryness Index (DI) in the black land Prairie of Eastern Mississippi. In the District of Gorantiwar, Meshram, & Mittal, (2011), used ARIMA models to model weekly ET_0 in semi-arid regions by using seasonal ARIMA) (1,1,0) (1,0,1) models. The results of the study proved that ARIMA was also feasible for predicting the reference crop evapo-transpiration, as it had the least value of RMSE when compared with actual ET_0 and prediction ET_0.

Birylo, Rzepeck, Kuczynska-Siehien, & Nastula, (2018), used the ARIMA model to assess the water ground in Suwalki, Zegrynski and Tarnow, three towns located in Europe and Poland. Drinking water is part of agriculture and, according to the European Union Water Framework Directive (EUWFD), efficient monitoring of water levels by each country in Europe is crucial and they normally use water budget to assess the accuracy of their determination of groundwater levels. This was done by studying three components of the water budget which are: precipitation, evapo-transpiration, and surface run-off. Three variables from January 1979 to July 2016 were computed and their results were compared with the results of ARIMA prediction over 12 months from August 2015 to August 2016. It was proved that the predicted results were correct, hence, there were higher cross-correlation values between the actual values and the predicted values.

In a study by Valipour, (2012), in Tehran, a town in Iran selected the Mehrabad Synoptic Station to test the ability of Box-Jenkins models, which use Auto-Regressive Integrated Moving Average (ARIMA), Auto-Regressive Moving Average (ARMA), Auto-Regressive (AR) and Moving Average (MA) models the predict the reference potential evapo-transpiration. The purpose of the study was to predict ET and other meteorological data by using Box-Jenkins models and to increase the number of parameters in order to expand the accuracy of the prediction of five selected variables and comparing their root mean square errors. (Valipour, 2012), collected the data for 45 years, from 1951 to 2000, from Mehrad Synoptic Station, for wind speed (U), mean temperature (T_{mean}), Maximum temperature (T_{max}), minimum temperature (T_{min}), dew point temperature (T_d) and sunshine hours recorded (n). Eight evapo-transpiration equations – FAO Penman Monteith (FPM), FAO Blaney Criddle (FBC), Turc, FAO Radiation Making (FRM), Priestley Taylor (PT), Hargreaves Samani (HS), Thornthwaite (TW), and Corrected Fanses Haise (CJH), and Box-Jetkins Method were used in the study to model the reference potential evapo-transpiration individually, using MINITAB software to run all the models. After running more structures using MINITAB, (Valipour, 2012), was able to compare the RMSE of all the Box-Jenkins models (ARIMA, ARMA, and AR). It was clear that these models were suitable tools to predict meteorological data and reference evapo-transpiration, but ARIMA models excelled as its RMSE was lower than the other models.

In a study by Bari, Rahman, Hussain, & Ray, (2015), in Sylhet-3114 in Bangladesh, ARIMA models were used to predict long-term rainfall for the years 1980 to 2010. The future prediction results with a 95% confidence interval, with the most efficient being ARIMA (0, 0, 1) $(1, 1, 1)_{12}$, were archived in the hope that the results would assist decision-maker's ineffective scheduling of flood prediction, rainwater collection, crop management, and urban planning. After the model was selected and tested, it was used to predict the data from the year 2007 to the year 2012, and those were compared with a test from the year 2007 to 2010 with a 95% confidence level, which gave a reasonable result proving that the model could be used in future studies around the area of Sylhet City.

In a study conducted in the Nile Delta region in Egypt, Psilovikos & Elhang, (2013) used ARIMA to predict evapo-transpiration for effective planning and management of water resources, as the environment was considered to be fragile. Data from the European Space Agency (ESA) were used to predict daily ET. AATSR data from the year 2005 to 2009, with a spatial resolution of 1 km at nadir, were used to estimate ET. The results that were predicted using ARIMA without referring to the seasonality of the data were unreliable, and iterations of the model input led to the model being inadequate for predicting as the accuracy was too low. Accurate prediction was done after the ARIMA model was refined, with six legs of seasonality being the most accurate and efficient. The final results of the combined Remote Sensing data and use of seasonal ARIMA models to predict the daily ET provided the base for enhanced and balanced control water resources in arid ecosystems matching the results obtained before. The study proved that increase in temperatures and ET lead to increased demand for agricultural water use in Egypt and the results led to a change in how water was supplied in the Nile Delta region.

ARIMA models were also used in the Jordan Valley, Hamdi, Bdour, & Tarawneh, (2008), because of its environmental challenge of scarcity of water, which has been referenced by the average of the previously estimated ET, which was over 90% of the country's total population. The study of reference evapo-transpiration was unique, using time-series analysis. Data from Central Jordan Valley for 29 years, from 1973 to 2002, were taken and pan-evapo-transpiration was used to reference the evapo-transpiration. After the model was evaluated and identified, and the reference evapo-transpiration was predicted using ARIMA (1, 0, 0) (0, 1,1), the model was developed, and it was able to provide acceptable prediction when it was compared with the results of the reference evapo-transpiration measured using the pan-evapo-transpiration parameter. Twelve different ARIMA models were tested and only three demonstrated acceptable performances when comparing their Root Mean Square Error (RMSE), Mean Absolute Forecast Percentage Error (MAFPE), and Maximum Absolute Forecast Error (MAFE). This study is believed to be crucial to Jordan as the town is ranked the lowest in water resources and the study area constituted 20% of the water scarcity. The developed ARIMA models are believed to be able to assist water managers as they enable them to predict up to five years with an error of only 11%. The findings were good enough to enable planning, design, and proper management of any irrigation scheme in that region, and they could also be used by policy-makers for effective management of crucial resources for irrigation water in a water-stressed country such as Jordan.

ARIMA models can also be used in semi-arid climate conditions. In a study by Gorantiwar, Meshram, & Mittal, (2011), in Solapur District, located in Maharashtra, reference evapo-transpiration was modelled by using ARIMA models. Data for 23 years, from 1984 to 2006, measured using the Penman Monteith method, were collected and different ARIMA models were used for prediction. There were 10 identified models and the best model results were selected based on the least error. Using the best ARIMA (1, 1, 0) (1, 0, 1) model selected based on the Root Mean Square Error (RMSE), the weekly reference crop evapo-transpiration for the year 2007 was predicted. It was proved that ARIMA models have the potential to predict the reference crop evapo-transpiration and they were able to produce the best stochastic model for creating and predicting the weekly ET_r values for the identified study area in India.

A similar study of crop reference prediction was conducted in the Bokaro District, located in Jharkhand, in India (Gautam & Sinha, 2016). The purpose of the study was to predict the Crop Reference Evapo-transpiration to provide clear information for optimal control of water resources, and ARIMA models were used for predicting the mean monthly reference crop evapo-transpiration based on stochastic analysis. Using 1224 data for the study, ARIMA (0, 1, 4) $(0, 1, 1)_{12}$ model was identified as the best model because the ACF and PACF residuals were significantly close to zero at 95% confidence level. Using the selected best model, a prediction of 24 months was carried out, and the Root Mean Square value indicated a very strong correlation between the observed data and the actual data. It was proven that the model developed could assist the water managers to formulate strategies on how to manage water and the model presented could also be used to predict more months, even though the accuracy would be compromised.

ARIMA models have been used widely in predicting evapo-transpiration over the years, using time-series data. To predict the impact of climate change in India, Sarangi, (Kambale, Singh, & Sarangi, 2017), used the ARIMA models to assess the trend and predict the changes in climate variables. In the study conducted in Raichur, the time-series data for 35 years were obtained from meteorological observations located in the National Capital Territory (NCT) of Delhi and were used to predict the ET and crop requirements. ARIMA was used to predict the future climate parameters until 2040 and, using the coefficient of determination, ARIMA models were able to give reliable information with regard to the decline in crop water requirements by 2030.

Bouznad, et al., (2020), used ARIMA to predict precipitation, temperature and evapo-transpiration in the Algeria Highlands. Monthly rainfall and temperature data for 30 years, from 1985 to 2014, were collected from 11 weather stations managed by the Algeria National Weather Office. The ARIMA model was built using a dataset of 360 monthly observations. Based on the study, it was possible to indicate areas which are affected by drought and ARIMA models were one of the models used.

Simons & Laryea, (2005), also used ARIMA to assess the efficiency of African markets between four countries, namely, Ghana, Egypt, Mauritius and South Africa. The models were used with other naïve models which they out-performed. The weekly and monthly data for 13 years, from 1990 to 2003, were constructed for South Africa, Egypt and Ghana, as well as price series on closing values every Wednesday. The Box-Jenkins ARIMA models successfully predicted the return that would be generated in Ghana, Egypt and Mauritius to exploit market inefficiency, and such predictions out-performed the naïve models also used on their study. To measure the trend of energy consumption variations in South Africa, (Ma & Wang, 2019), used ARIMA, NGM and NGM-ARIMA models. The energy consumption data for South Africa were obtained from the statistical yearbook of BP energy for 18 years from 1998 to 2016. Using MAPE statistical model measuring tools, the results indicated that the hybrid NGM-ARIMA model was highly accurate in predicting the energy consumption in South Africa. In South Africa, ARIMA models have been used in various fields, using the available time-series data. Saayman & Saayman, (2008), used ARIMA models in forecasting the tourist arrivals in major inter-continental tourism markets in South Africa. Saayman & Saayman, (2008) used data showing monthly arrivals of tourists for 12 years, from 1994 to 2006. The time series data used for the study were collected from Statistics South Africa and proved that there was a significant growth in South Africa's tourism over this period. The ARIMA models proved to be the best predictor and performed accurately over three time horizons, which were three months, six months and twelve months.

Wadi, Ismail, Alkhahazaleh, & Abdul Karim, (2011), used the wavelet transform models to predict financial time series data based on ARIMA models. The daily return time-series dataset for April 1993 to December 2009 consisted of 4096 observations. The wavelet transform models performed better than the original data used. Wadi, Ismail, Alkhahazaleh, & Abdul Karim, (2011) also concluded that results forecast by ARIMA indicated more accuracy regarding the financial stock market.

To predict the demand for food, historical data were used by (Fattah, Ezzine, Aman, Moussami, & Lachhab, 2018). Data collected from January 2010 to December 2015 were used to develop and validate various ARIMA models before predicting demand in order to improve the planning process that could be used in the future by businesses. Using four performance criteria, the best ARIMA (1, 0, 1) was selected, and predictions proved that the model is a perfect tool that can be used by manufacturing companies in making significant decisions because of its reliability. The model has also been used by Chen, Yuan, & Shu, (2008), to predict property crime in China over 50 weeks. Chen, Yuan, & Shu, (2008), used ARIMA models, fits and predictions compared with other models in order to assist police stations and municipal officials in making informed decisions to prevent crime. The data were collected from 110 computer-aided dispatch (CAD) records from the local police station for crimes including theft, robbery and burglary which accounted for 90% of crimes in that region. ARIMA models performed better than Simple Exponential Smoothing (SES) and Holt two-parameter Exponential Smoothing (HES) models respectively.

The models have also been used in India to predict the stock prices (Mondal, Shit, & Goswami, 2014), used National Stock Exchange (NES) data for 56 companies from 7 sectors, and 8 firms from the official website of the India National Stock Exchange. The data from April 2012 to February 2014 were used to predict prices for future months from September 2013 to February 2014, and data for the other eight companies was used to predict prices from September 2012 to February 2014. The efficiency of the ARIMA models was proved beyond any reasonable doubt with accuracy above 85% in predicting the prices of stock.

2.10.2 Artificial Neural Networks (ANNs)

ANNs have been used in various fields for predicting and modelling. ANNs were adopted for the current study because of their ability to predict non-linear time series. Many researchers from different areas of the world have used ANNs before and archived superior results.

Tawegoun, Belbrahem, & Chasseriaux, (2004) used ANNs to predict ET on a nursery area for irrigation management purposes. The data for 14 days, from 21 July to 3 August, which consisted of four climate databases, including global radiation, relative humidity, air temperature, and wind speed, were collected from the experimental centre of the nursery area of the National Agronomic Research Centre of Angiers. To achieve an accurate neural network model, two ANN models were selected using Root Mean Square Error. The first test was for model speed and memory. After training, the feedback was capture and the behaviour of the model was described as accurate, and it was concluded that ANNs could be used in a context in which irrigation management was anticipated. Hence, the simulated prediction using steady data showed how adequately the models predicted ET.

In a study by Traore, Wang, & Kerh, (2008), conducted in Burkina Faso, Artificial Neural Networks were used to predict reference evapo-transpiration. A generalised regression neural network (GRNN) model was selected because of its ability to predict even if there is not enough climate data, hence the normal Penman-Monteith (PM) equation could not be used. A GRNN model is also favoured in place of a multi-layer network because it does not need any training. Owing to the limited data, a reference model for Bukina Faso (RMBF) was developed using the temperature for intended irrigation management in Banfora and Ougadougou. Four models, GRNN, EMBF, Hargreaves (HRG), and Blaney-Criddle (BCR). were used to investigate the performance of the model in Dori, Bogande, and Fada N’gourma, using the data collected in the Dedougou region from 1996 to 2006. The data were divided into three parts: data from 1996 to 2003 mainly for training the model, data from 2004 to 2005 were used for cross authentication of the model, and data from 2005 to 2006 were used for testing. The study proved that GRNN can be used in areas with limited climate data, and it provided more accurate results than other methods employed, on condition that the wind velocity is considered in the model. Furthermore, the study proved that, in the absence of climate data. ANNs can be used in semi-arid regions in Africa.

ANNs were employed by Pandorfi, et al., (2016), to predict the evapo-transpiration of sweet pepper cultivated in a greenhouse. Meteorological data for 135 days were collected from a simple “arched-roof East-West structure with a length of 17.5 m x 6.4 m x 3.0 m with arch height of 1.2 m enclosed with 0.15 mm thick low-density polyethylene (LDPE) film”. The data from September 2013 to February 2014 included temperature, wind speed, relative air humidity, solar radiation, and evapo-transpiration, which were determined using data obtained by load cell weighing lysimeter. Neural networks were measured using a devoted computer programme and the data used were divided into three sets: 40% for training, 20% for testing, and 40% for validation. An error back-propagation model was used in the study to develop the neural network. It was possible to establish the evapo-transpiration pattern through the application of ANNs even though the accuracy of the performance of the system had to be monitored regularly so that the network could be maintained. The success of the ANN model in the study was associated with its ability to adapt, which led to them being considered as a promising model for decision-making. The purpose of the study which was to test the viability of using ANNs to predict the evapo-transpiration of sweet pepper in a protected environment was archived. The results were compared with the results of the weighing lysimetric method and ANN results proved to be good.

In a study by Abdullahi & Elkiran, (2018), the competence was determined of ANNs, trained using a three-layered network trained by Feed Forward Back Propagation and Levenberg-Marquardt, to predict monthly reference (ET_0) using data for 408 months. The data obtained from the National Aeronautics and Space Administration (NASA) from January 1983 to December 2016 were used to study ET_0 at Kyrenia/Girne, located in Northern Cyprus, and Larnaca, located in Southern Cyprus. To avoid model computational overloading, model overfitting, and under-fitting, ANNs were trained using FFBP and LM optimisation algorithms, which was done by first keeping input parameters constant and changing the number of hidden layers for twelve trials. In the second approach, input parameters were changed consecutively with hidden layers doubling the parameters, and a model structure was selected. To test both approaches through trial and error, the best model structure was selected based on lower Root

Mean Square Error (RMSE) and Higher Determination Coefficient (R^2). ANN proved to be effective in predicting the impact of climate on reference evapo-transpiration, even when the model is used with limited data. The conclusions showed that ANN could be trusted as a reliable tool to predict future global warming effects in Cyprus Island.

Kumar, Raghuwanshi, SIngh, Wallender, & Pruitt, (2002), used ANNs to predict daily grass reference crop evapo-transpiration (ET_0) and the model performance was compared with the traditional Penman-Monteith method. The daily climate data (maximum temperatures, minimum temperatures, relative humidity, solar radiation, and wind speed) were collected from the Davis California Irrigation Management Information System Station for the period from 1 January 1990 to June 2000. The daily ET_0 was predicted using the traditional, standard PM method and these values were used to develop the ANN model. To obtain the most promising model, ET_0 was determined using a 108, single hidden layer network and the results were compared with results predicted by PM and the best solo learning technique was chosen based on the architecture of the network and Weighted Standard Error of Estimate (WSEE). The results of the study proved that one hidden layer ANNs are adequate to account for the nonlinear connection between climate variables and ET_0. The results predicted for ET_0 suggested that ANNs were better than the standard PM method, as all the ANN trained models produced lower WSEE than PM.

Tawegoun, Belbrahem, & Chasseriaux, (2004), used neural networks to predict evapo-transpiration in a nursery. The prediction of the change in local climate conditions is very important for the survival of plants. In the study data (global radiation, relative humidity, air temperature, and wind speed) for 14 days from 21 July 1994 to 03 August 1994 were collected from an experimental nursery at the National Agronomic Research Centre of Angiers. The study showed that evapo-transpiration could be predicted using steady-state data by using a Recurrent Neural Network. The results showed that the model can be used for simulation in a context in which irrigation management is anticipated.

For reliability of irrigation management Kishore & Pushpalatha, (2017), used Artificial Neural Networks to compare the results of the reference evapo-transpiration (ET_0) in

Kanchipuram District in Tamil Nadu, India. Two models were used based on their common use for predicting time series and ANN models. To avoid the model overfitting, the Lavenberg Marquardt (LM) back-propagation training algorithm was used. LM adjusted all the bias values and trained the network for n = 100. The results study proved that ARIMA models can be used accurately for short-term prediction, whilst ANN can be used for long-term data. Kishore & Pushpalatha, (2017), suggested that future research can be done on prediction using machine learning.

2.10.3 Hybrid (ARIMA-ANN) Model

The Hybrid (ARIMA-ANN) model is one of the models used to predict evapo-transpiration at the Keiskammahoek Irrigation Scheme. Owing to the complexity involved in determining the linearity and non-linearity of the time series and selecting the most effective model to use for prediction, both ARIMA and ANN models are used, as their uniqueness is based on conditions. Using several different models means that the best model is selected based on the accuracy of results, while obtaining adequate results from one model might not necessarily prove that specific model to be accurate, because conditions might favour that model. Using both models makes it easier to select the ideal model, also because purely linear or non-linear problems do not always exist in the real world (Zhang G. P., 2003).

Other researchers have also suggested the use of different models when modelling, as there is no perfect model (Makridakis, et al., 1982). It is clear that there is no universal, adequate model to choose between ARIMA and ANN that is suitable for all timeseries since real-world time series comprise both linear and non-linear relationship structures amongst the observations (Khandelwal, Adhikari, & Verma, 2015). Soh, Koo, Huang, & Fung, (2018), suggested that the combined models combine the strength of different models, thus giving many accurate results.

2.10.3.1 Use of different hybrid models

Zhang G. P., (2003) used hybrid models to check the accuracy of Auto-Regressive Integrated Moving Average and Artificial Neural Networks. The aim of the study was to take advantage of the unique strength of both models, as ARIMA predicts linearity and ANN predicts non-linearity. The results proved that using a combined model significantly reduced the error, even though ANN was better in some periods. Zhang G. P., (2003) suggested that the use of a hybrid model using linear and non-linear data is ideal for complex problems, and the model can be used to improve the performance of prediction.

To check the accuracy of the hybrid model Khandelwal, Adhikari, & Verma, (2015), used three different models: the Discrete Wavelet Transform to decompose the training set of linear and non-linear data and then used ARIMA and ANN to identify and predict the reconstructed, complete, and estimated components individually. Therefore, the selected approach used the unique strengths of DWT, ARIMA, and ANN to improve the accuracy of prediction. Using Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) values it could be concluded that Hybrid (ARIMA-ANN) models performed more accurately compared with single models, even though the proposed model that applies both ARIMA and ANN to model linear and non-linear components, after decomposition using the DWT model, performed much better than all three models, Hybrid (ARIMA-ANN), ANN and ARIMA.

Hybrid (ARIMA-ANN) models can also be used in other fields such as stock-market prediction. Elwasify, (2015), searched for the most accurate model to use to model the EDX 30 stock market index. Data containing 270 daily observations, in the form of a time series, were used from 1 October 2009 to 31 October 2010. The Multi-layer Perceptron ANN was used, which has three layers: input layer, hidden; layer, and an output layer based on ANN (1:6:1) and ARIMA (0,2,2). Mean Absolute Relative Prediction Error (MAPE) and Mean Square Relative Predictions Error (MSRE) were used to compare model performance. The Hybrid (ARIMA-ANN) performed very well with error levels lower than when each model was used alone.

To predict and analyse water quality, Zhang, Zhang, & Li, (2016) used a Hybrid Auto-Regressive Integrated Moving Average (ARIMA) and Radial Basis Function Neural Network (RBFNN) to estimate and analyse the water quality at Chagan Lake. ARIMA, based on Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF), was selected as ARIMA (2.1.1). Prediction residuals from ARIMA (1,1,1) were used to input RBFNN cells and a Hybrid (ARIMA-RBFNN) model was developed. Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) were used, and the results showed that the Hybrid model gave many accurate results compared with ARIMA and RBFNN models which had lower RMSE and MAPE values. Zhang, Zhang, & Li, (2016) concluded that hybrid models can be used, and the results served as the theoretical basis for environmental management of Chagan Lake.

Diaz-Robles, et al., (2008) used Hybrid (ARIMA-ANN) to improve the prediction of air quality time series in Tumco, located in Chile. Two models, ARIMA and Multilinear Regression (MLR) were used but, because of their inability to model extreme events, a combined Novel Hybrid model had to be used. Data from Temco Station for 6 years, from 200 to 2006, were used. ARIMA (p,0,q) together with predator variable X, was then called ARIMAX or MARIMA. Independent variables, X, were used to compose the MLT, and ARIMA was attained by using the time-series predicting tool called SAS 9.1 Software. To develop ANN, the multi-layer perceptron (MLP) ANN was used to form model with Levenberg Marquardt (LM) and, using the Enterprise Miner tool of SAS 9.1, a training algorithm was used to form the non-linear version. To form the Hybrid (ARIMA-ANN) model, two steps were followed: firstly, to develop the ARIMAX model to predict Max PM and, secondly, to develop ARIMA to define the results obtained from the ARIMAX model, and the Enterprise Miner tool was used to develop the Hybrid model. The results of the study proved that the Hybrid (ARIMA-ANN) model was the best, giving more accurate results than other models, based on its exclusive abilities to model linear and non-linear data.

The Hybrid (ARIMA-ANN) model can also be used to improve the competitiveness of the Turkish Power Market, and the proposed model gave results with few errors. The same process was followed of developing ARIMA using Box-Jenkins with three steps followed by creation of ANN, using residuals from ARIMA. Model accuracy was checked using Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE). The model performed very well, and it was proposed that other markets could also use it. Use of the Hybrid model in developing countries was further recommended by the energy investors to obtain accurate results.

2.10.3.2 Averaged models

In the current study, averaged ARIMA, ANN and Hybrid (ARIMA-ANN) models were used. This was done to investigate whether there would be any accuracy or promise that averaged models would be fit to make ET predictions for the Keiskammahoek Irrigation Scheme. The averaged ET data from three previous models were then used. Several model averaging studies are gaining much attention in literature about hydrology, atmosphere, and statistics in order to address uncertainty of model results. High-density prediction is done in order to assist in analysing some parts of models that might have uncertainties (Diks & Vrugt, 2010).

Diks & Vrugt, (2010), used averaged models and compared their applications in hydrology. Two sets of data from Leaf River watershed in Mississippi, United States of America were used, which consisted of daily streamflow data for 36 years recorded historically. The hydrological data in this classical prediction scenario included rainfall run-off and water hydrology. Eight widely used, conceptual, hydrologic models were used to predict the dynamics of the streamflow. A second dataset included soil hydrology and the flow of soil water through vadose areas was predicted using a dataset from layered vadose areas in New Zealand. One of the conclusions from the study was that the model performed well based on RMSE model performance measurements.

In a review study by Kim, Jeong, & Ko, (2006), various models which have been used to make predictions in economics were assessed and their ability in hydrology was evaluated. Different models were used, including simple average, Sum of Squared Error, Artificial Neural Networks, alternating regression and constant coefficient regression methods. In all of these models, streamflow forecasts were combined using the existing model for rainfall run-off. Other models were developed using new rainfall run-off models that used assembled Neural Networks for predicting the monthly inflow of the Korea Dam, called Daecheong Multi-purpose. The combined models out-performed other models, based on Sum of Squared Error, and enhanced the probabilistic prediction accuracy of the known WSP systems.

Granger, (1984), also used the model combination to assess the accuracy of their study in which three approaches were used to obtain linear combinations. The method of getting averaged weighted predictions was not applied, but three techniques were considered in which constant terms were added without constraining the weights in order to achieve unity. The approaches were tested with predictions of hog prices quarterly both within the sample and outside the sample. It was found that the method of combining these models out-performed common practice.

In another study by Hoeting, Madigan, Raftery, & Volinsky, (1999), Bayesian Model Averaging (BMA) was used to reduce the model uncertainties. Volinsky *et al.* 1999 stated that the generic method used by data analysts generally selected the model from various data and continued assuming that a chosen model generated the data. Hoeting, Madigan, Raftery, & Volinsky, (1999), stated further that these models do not take model selection into consideration, and this leads to over-confidence resulting in risky inferences and decisions. Hoeting, Madigan, Raftery, & Volinsky, (1999), used BMA, arguing that it includes coherent techniques for making allowance for uncertainties in models. Volinsky *et al.* 1999 proved that BMA produces improved prediction results.

2.11 R AND R-STUDIO

R is an open-source package and the statistical technique that provides facility for visual and data analytics. It also provides strong graphical support and, recently, for forecasting analytics to arrive at better decisions and finding solutions. Various industries have used the R tool in making efficient decisions and their businesses have improved. The tool has features such as being an open software package, it uses simple and efficient programming language, it is effective for visual analytics, the programme can withstand huge amounts of data, and it is efficient in manipulation and storage (Shinde, Oza, & Kamat, 2017). According to (Freeman & Moisen, 2008), the tool has a set of functions which are important in assessing the outcome of the analysis of available and absent data. The R tool includes toolkit which makes it possible to select the ideal threshold for transforming a probability surface into available and absent data maps which are tailored to a particular intended use.

R-Studio is also an open-source software package that collects and runs on diverse operating systems for arithmetical computation and production of graphics. The programme was initiated by Robert Gentleman and Ross Ihaka and the language it uses was greatly influenced by S Language, developed by John Chambers and Colleagues at Bell Laboratories. S Language has become the *lingua franca* for statistical computation in various fields of trade and research. It is premised on its main scripting language, but it permits blending with compiled code transcribed in Java, C, C++ and Fortran for rigorous computational tasks or for applying kits stipulated for added languages (Verzani, 2002). It is a free programming language used mostly for statistical computation by statisticians and data miners based on a broad dimension of pattern withdrawal and the language offers various statistical methods from classification to gathering and examination (Hussain, 2015).

CHAPTER 3: MATERIAL AND METHODS

3.1 DESCRIPTION OF THE STUDY SITE

Keiskammahoek Irrigation Scheme is one of the schemes developed in the old “Bantustans/Native areas” during the 1950s and 1960s by the Ciskei Government (Figure 3.1). The project of developing irrigation schemes in homelands was triggered by the Tomlison Commission recommendations after the development of homelands administration during the 1970s when irrigation was found to be the most favourable way to develop the Ciskei area (Averbeke, M'Marete, Igodan, & Belete, 1998).

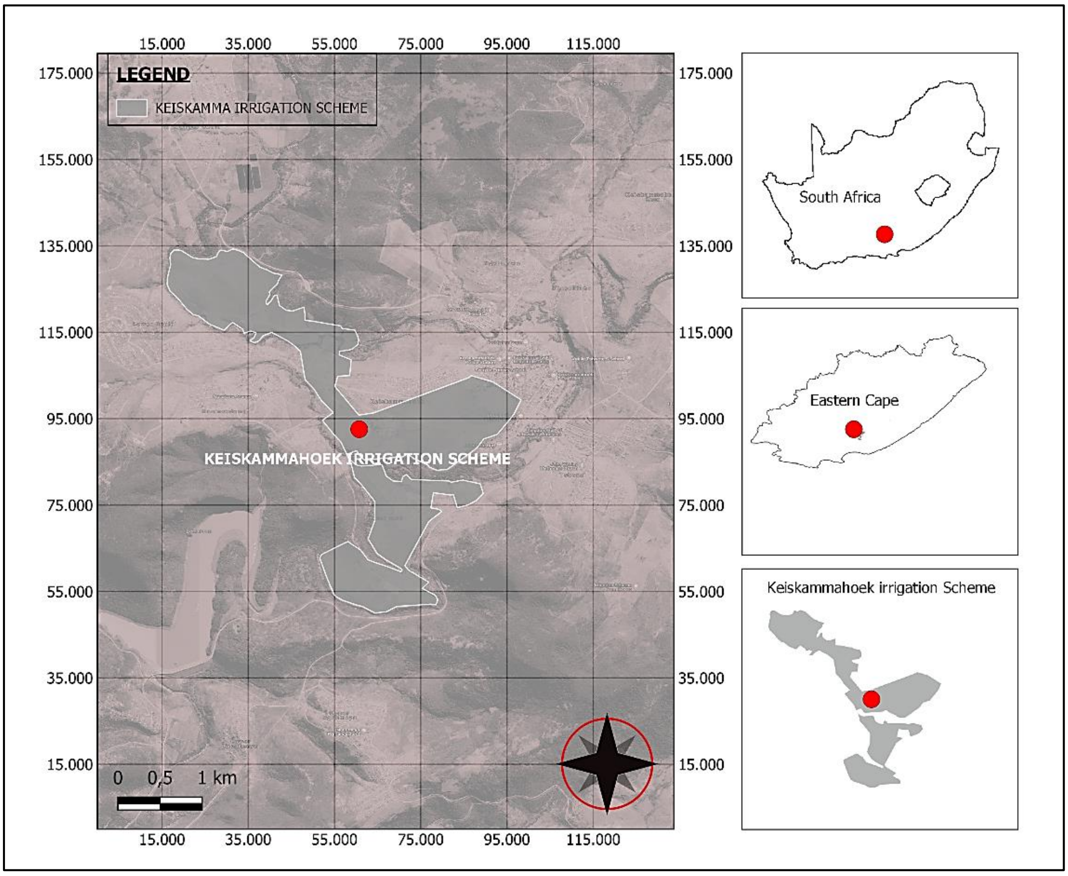


Figure 3.1: Keiskammahoek Irrigation Scheme

[Source: Q.GIS]

The study site is situated in the west of Keiskammahoek Town in Amahlathi Local Municipality in the district of Amathole in Eastern Cape, South Africa. The site is 40 km west of King Williams Town and 36 km South-West of a small town called Stutterheim. The geographical location of Keiskammahoek Irrigation Scheme is Latitude S 32°41'14" E 27°07'48". The area temperature in King Williams Town, which is the closest area to the Keiskammahoek Irrigation Scheme, ranges from 6.5º C in winter to 26.7º C in summer. King Williams Town receives an average annual rainfall of 502 mm in summer (Affari, Department of Environmental, GIBB, & South African National Road Agency SOC ltd, 2016). According to Brown, (1969), the region have humid temperature climate with sufficient rainfall in the season, however dry in winter with a mean annual temperature if the warmest month being below 71. 60. Keiskammahoek is located in Eastern Cape Province and according to (Apraku, Akpan, & Moyo, 2019), the province is highly exposed to climate change impact whilst it is also leading with least adaptive measures due to the lack of knowledge.

Figure 3.2 shows the historical time-series data of Keiskammahoek Irrigation Scheme, with Figure 3.2 (a) indicating the time-series data of the Latent Heat Flux Effect and Figure 3.2 (b) indicating the data for precipitation for the 18 years of the study period. These two indices were also collected at Keiskammahoek Irrigation Scheme using Java script to extract from the Modern-Era Retrospective (MERRA-2) model. The use of this model in the hydrological cycle was motivated by the inadequacy of other models to analyse the weather and climate studies (Rienecker, et al., 2011). The model data are obtainable at 0.67º x 0.50º resolution at 1 to 6 h intervals (Xulu, Peerbhay, Gebreslaie, & Ismail, 2018). This 18 years, monthly averaged data from 2001 to 2018 of Latent heat flux and Precipitation is very important as these indices give an idea of how the climate behaviour is at keiskammahoek Irrigation Scheme.

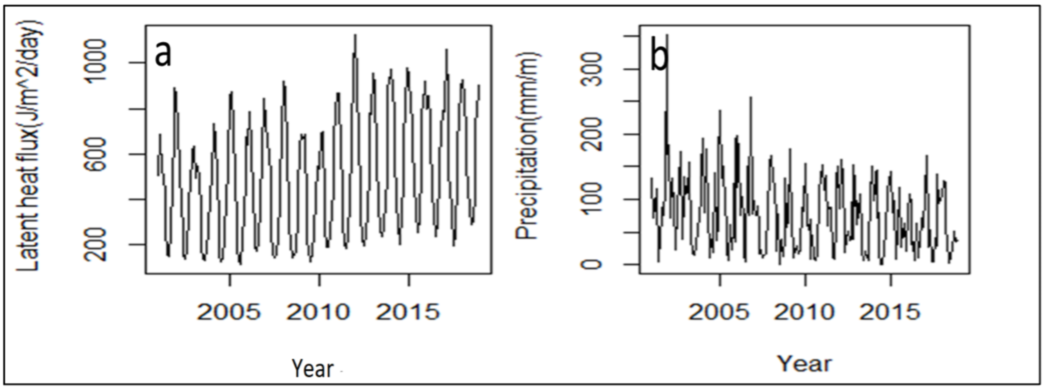


Figure 3.2: ALHF and precipitation of Keiskammahoek Irrigation Scheme.

[Source: created by R]

3.2 RESEARCH DESIGN

A quantitative method was used for the study. “Quantitative research is the explaining of phenomena by collecting numerical data that are analysed using mathematically based methods (in particular statistics)” (Bhawna & Gobind, 2015). It is the method by which numeric data are gathered to clarify a particular situation (Muijs, 2004).

The following flow chart shows the research design.

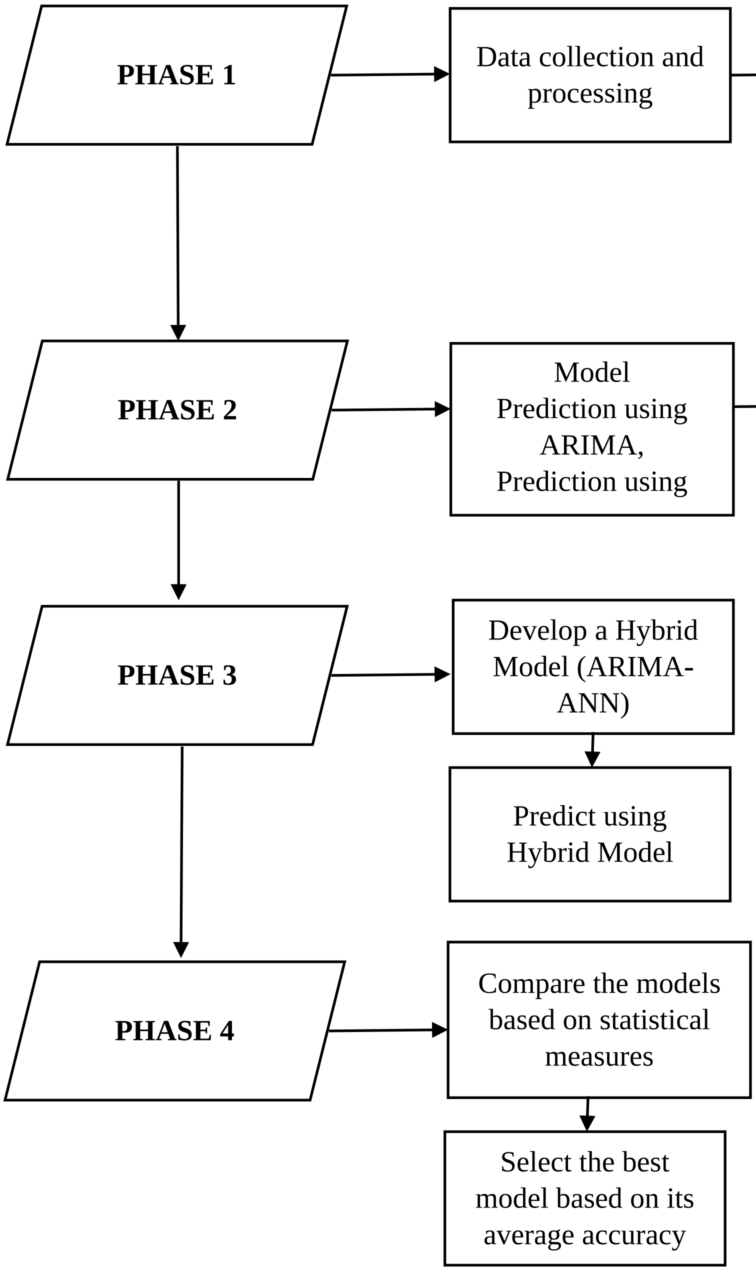


Figure 3.3: A flow chart of the research design

Figure 3.3 Shows how the research is designed. The research is designed in four phases with phase 1 being the Data collection, ARIMA development and Model calibration. This is where the objective one of this study will be meet. The available ET data from the study site will be collected and ARIMA model will be calibrated in order to ensure that accurate prediction will take place. The second phase will be the modelling of ET using two modelling techniques which are ARIMA and ANNS and these results are compared. The phase 3 of this study shall deal with Developing a Hybrid (ARIMA-ANN) models and predicting using this model. And lastly stage 4 will be comparing the three prediction and select the best model that is promising for estimating ET for Keiskammahork Irrigation Scheme.

3.3 DATA COLLECTION

In this study, the ET time-series data collected for the period of 18 years (2001-2018), using the satellite remote sensing, were used. Satellite remote sensing offers the essential means to retrieve this data at flexible intervals (Xulu, Peerbhay, Gebreslaie, & Ismail, 2018). Other parameters which are linked to ET were also considered to assess the drought possibility in the study area. These were: the Normalised Difference Water Index (NDWI), which is capable of defining the water features and improving identification of the existence of water in remotely sensed, digital imagery (Maselli, Rodolfi, Botta, & Conese, 1996); the Normalised Difference Vegetation Index (NDVI), which was also analysed to quantify the amount of green vegetation in a region (Sultana, et al., 2014); the Normalised Difference Drought Index (NDDI), which is the combination of NDVI and NDWI, which was used specifically to monitor drought (Du, Bui, Nguyen, & Lee, 2018). Finally, precipitation (P) was also used to check the relationship between ET and precipitation. A portion of the ET time-series data was used to train the model, while the remaining data were used to predict ET, using the Auto-Regressive Integrated Moving Average model (ARIMA), Artificial Neural Networks (ANN) model and Hybrid models.

Three sources were used to collect data for this study. The Normalized Difference Vegetation Index (NDVI) and the Normalised Difference Water Index (NDWI) were used to obtain the current ET in Keiskammahoek Irrigation Scheme. Data for 18 years from 2001 to 2018 were gathered using Java script, the monthly averaged MODIS

Terra/Aqua 16-day dataset. (MOD16A2) is the cloud-built software podium for geospatial examination on a universal scale that conveys the considerable computational abilities of Google to observe a variety of high effect societal matters comprising several variables such as “deforestation, drought, disaster, disease, food security, water management, climate monitoring and environmental protection (Gorelick, et al., 2017). Figure 3.2 depicts the Google Earth Enginer interative development environment. The orange overlay shows the Keiskammahoek iriigation scheme site and the ET time series, which is averaged for the study area from 2001 to 2018 as shows on the console part pf the GEE interactive development environment.

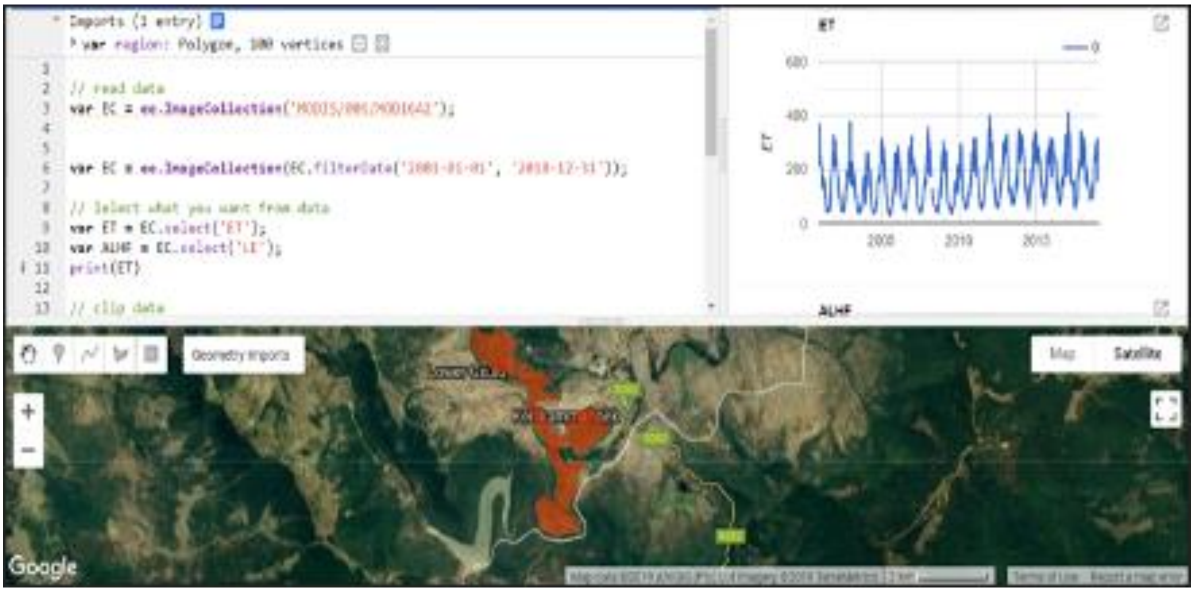


Figure 3.4: Extracted Keiskammahoek Irrigation Scheme

[Source: MOD16A3.006, 2001]

“The MOD16A2 evapo-transpiration product is an 8-day composite product produces at 500-meter pixel resolution” (MOD16A3.006, 2001). The MOD16 ET algorithm is founded on the substantially sound concept of the Penman-Monteith energy balance (Jovanovic *et al*., 2015). Another variable that was used in this study was precipitation, which was extracted from the Modern-Era Retrospective (MERRA-2) model. The use of this model in the hydrological cycle was motivated by the inadequacy of other models to analyse the weather and climate studies (Rienecker, et al., 2011).The model data are obtainable at 0.67^{o} x 0.50^{o} resolution at 1 to 6 h intervals (Xulu, Peerbhay, Gebreslaie, & Ismail, 2018). Normalised Different Vegetation Index (NDVI), Normalised Difference Vegetation Water Index (NDWI) and Normalised Difference Drought Index calculations were defined by (Tavazohi & Nadoushan, 2018) as follows:

$$NDWI = \frac{\rho NIR - \rho SNIR}{\rho NIR + \rho SNIR} \tag{3.1}$$

$$NDVI = \frac{\rho NIR - \rho Red}{\rho NIR + \rho Red} \tag{3.2}$$

Where ρNIR is the near infrared band and ρSNIR is the short infrared band. NDDI is represented by a combination of NDVI and NDWI in the following formula:

$$NDDI = \frac{NDVI - NDWI}{NDVI + NDWI} \tag{3.3}$$

The monthly Stream Flow (SF) and monthly volumes Sandile Dam were collected from the Department of Water and Sanitation (DWS) website with Stream Flow collected from Station R1H015-RIV. The purpose of using Streamflow of the Keiskammahoek River is to assess the volume of water on this river over the study period. This will assist in evaluating whether the supply river to Keiskammahoek Irrigation Scheme is under stressed over the years or not.

Figure 3.5: Sandile Dam (Source: AfriWX)

3.4 DATA ANALYSIS

After all the data was collected, different approaches to analysing statistical data were used to compare the evapo-transpiration with all the other parameters used. R-Studio programming software, which is excellent statistical analysing software, was used. The software is designed mainly for statistics and graphics, and it can also be used to collect data and produce presentation documents (Gandrud, 2018).

For data analysis, 3 models have been used which are ARIMA, ANNs and Hybrid (ARIMA-ANN). The main reason is that, as much as ARIMA is widely used, it still has disadvantages in its inability to predict non-linear time series. ANNs were then used as the second model because of their ability to predict non-linear time series. After the two models had been employed and evaluated, the Hybrid Model (ARIMA-ANNs) was employed to ensure that both linear and non-linear components of the time series were considered.

3.4.1 WAVELET ANALYSIS

In this study, Wavelet Analysis was used. Wavelet Transform was invented during the 1980s in geophysics mainly for examining seismic signals and their versatility, attractiveness, and wide use are the result of their exceptional properties (Zhang, Patuwo, & Hu, 1998). Wavelet Analysis was employed in this study because of the potential to obtain a time-frequency illustration of any continuous signal. The advantage of using Wavelet Analysis over other methods is that it can be used to examine the localised difference of power on the time series. “By decomposing a time-series into time-frequency space, one can determine the dominant of variability and their variation with time” (Loua, et al., 2019). There are various types of wavelets, but the choice of the purpose of the wavelet depends on the data series and the function of Morlet Wavelets for geophysical data has been proven as they have attained good results (Mbatha & Xulu, 2018). Therefore, the Continuous Wavelet Transform has CWT [W_n] for a specified time series ($X_{n,}$ n = 1, 2, 3……., n). (Torrence & Compo, 1997) explained the wavelet Ψ_0 (n) as:

$$\mathrm{W}_n(S) = \sum_{n'=0}^{n-1} X_{n'} \Psi * \left[\frac{(n'-n)\delta t}{s}\right] \tag{3.4}$$

Where Ψ is the complex conjugate of a wavelet, n' indicates the localised time index and s is the wavelet scale. This allows anyone to construct a photo demonstrating both amplitudes of any feature (evapo-transpiration in this case) versus the scale and how such amplitude changes with time. “The subscript 0 on Ψ has been dropped to indicate that this Ψ has also been normalized” (Torrence & Compo, 1997).

3.4.2 Analysis of Time Series Coherence Using Wavelet Coherence

Wavelet coherence was one of the time-series analysis methods used in this study. It is one of the techniques used to analyse the coherence and phase lag amongst the two time-series variables (Loua, et al., 2019); (Grinsted, Moore, & Jevrejeva, 2004) .The technique has gained much acceptance in time-series examination because of its ability to undertake the problem of continuous window width and concern of time resolution over the frequencies (NG & Chan, 2012). According to Loua, et al., (2019) the connection between two variables in this method is depicted by the arrows, which becomes the anti-phase relationship but, if the arrows point to the right, the two cross wavelet variables are in phase. This tool is defined by Grinsted, Moore, & Jevrejeva, (2004) with the following formula:

$$R_n^2(s) = \frac{\left|S(s^{-1} W_n^{XY}(s))\right|^2}{S(s^{-1}\left|W_n^X(s)\right|^2 . S(s^{-1}\left|W_n^Y(s)\right|^2)} \tag{3.5}$$

Where S represents the smoothing operator. It is noted that the above definition is similar to that of the traditional correlation coefficient. Therefore, it is important to think of the wavelet coherence as the confined correlation coefficient in time frequency. Therefore, the smoothing operator is written as follows:

$$S(W) = S_{scale}(S_{time}(W_n(s))) \tag{3.6}$$

Where S_{scale} means smoothing along the wavelet scale axis and S_{time} smoothing in time. It is essential to design the smoothing operator so that it can match the similar imprint as the wavelets used. According to Grinsted, Moore, & Jevrejeva, (2004), the suitable smoothing operator for Morlet Wavelet is given by the following formula:

$$S_{time}\,(W)|\mathrm{I}_s = \left(W_n(s) * c_1^{\frac{-t^2}{2s^2}}\right)\mathrm{I}_s \qquad (3.7)$$

$$S_{time}\,(W)|\mathrm{I}_s = (W_n(s) * c_2\,\Pi(0.6s))|_n \qquad (3.8)$$

Where c_1 and c_2 shows the normalisation constant and Π represents the rectangular function. The factor 0.6 is the empirically resolute scale decorrelation length for Morlet Wavelet.

3.4.3 Correlation Statistics Test Using Correlation Statistics

To assess the relationship between ET and other tested variables in this study, the Pearson Correlation Coefficient (PCC) was used. The PCC is a firm quantifier of correlation, which was originated by Galton in the late 19th century, who named it after his colleague. The coefficient has a range of + 1 (flawless correlation) to -1 (faultless but negative association) with 0 meaning the non-appearance of a relationship between the variables (Adler & Part, 2010). The PCC is given by the following formula:

$$r = \frac{\sum(R_i - R_{av})\,.(G_i - G_{av})}{\sqrt{\sum(R_i - R_{av})^2\,.\sum(R_i - R_{av})^2}} \qquad (3.9)$$

R_i is the strength of the primary red fluorophore in single pixels and R_{av}, is the arithmetic mean; G_i and G_{av} are the equivalent strengths for the second green fluorophore in the pixels.

3.4.4 Trend Slope Assessment Using Theil-Sen's Estimator

Theil-Sen slope estimator (TSE) was used in the analysis of long-term data and their seasonality of ET and other variables. TSE has proven to be strong, steady, and asymptotically ordinary under mild conditions and it is very proficient when the error circulation is constant, and it can be used to see the difference in robustness and effectiveness with the slightest squares predictions and to authorise super efficiency. The slope is very competitive compared with others because of its effectiveness in use and it has greater asymptotic competence with a high breakdown point of 29.3% (Dang, Peng, Wang, & Zhang, 2008). The following formula used in a study by (Loua, et al., 2019) was used in this study:

$$T_i = \frac{x_j - x_i}{j - i} \qquad (3.10)$$

Where x_j and x_i represent data values at time j and $i(j>i)$.

3.4.5 Trend Analysis Using Mann-Kendall Test

Trend examination has proven to be a crucial instrument for efficient water resource design and development, and subsequently running the trend discovery of hydrological properties, such as precipitation and stream flow, gives critical evidence on the likelihood of whether any propensity to change of a particular variable is imminent. Mann-Kendall tests are one of the most commonly used tests to notice trends in hydrologic time series (Yue & Wang, 2004). Machida, Andrzejak, Matias, & Vicente, (2013) explained the Mann-Kendall test (MK) as a “non-parametric test for monotonic trend detection in time-series data”. Spatial and temporal similarity of trends are crucial aspects of trend examination for any hydro-metrological limitations in various areas (Dinpashoh, Jhajharia, Fakheri-Fard, Singh, & Kahya, 2011). This non-parametric test has been used widely to examine the consequence of the trend in hydrological time sequences. To perform this test requires the testing data to be successively independent (Yue & Wang, 2004).

The MK test is non-parametric and, therefore, it is independent of missing data, the quantity of the data, irregular periods of monitoring the data, and the hypothesis of distribution, meaning it does not have to be a normal or bell-shaped distribution.

The assumptions of the MK Test are that:

- In the absence of a trend, the observations collected over time are not dependent and are correspondingly distributed.
- The attained observations are the precise representations of conditions during the time of sampling.
- The combination of the collection and its measurement methods offers reasonable and illustrative observations of the fundamental sample population over the studied time.

For this study, the Mann-Kendall test was applied to the evapo-transpiration using R-software under the construct of (Pohlert, 2016):

$$S = \sum_{k=1}^{n} \sum_{j=k+1}^{n} sng(X_j - X_k \quad (3.11)$$

Where $$sgn(x) = \begin{cases} 1 & if \quad x > 1 \\ 0 & if \quad x = 0 \\ -1 & if \quad x < 0 \end{cases} \quad (3.12)$$

The average value of S is E[S] = 0, and the variance σ^2 is given by the following equation:

$$\sigma^2 = \left\{ \frac{n(n-1)(2n+5) \sum_{i-1}^{p} t_1(t_1-1)(2t_1+5)dy}{18} \right\} \quad (3.13)$$

Where t_j is the number of data points in the *j*t tied group, and *p* is the number of the tied group in the time series. It is noted that the summation operator in the above equation is applied only in the case of tied groups in the time series to reduce the influence of individual values in tied groups in the ranked statistics. On the assumption of random and independent time series, the statistic S is approximately normally distributed provided that the following z-transformation equation is used:

$$Z = \begin{cases} \frac{s-1}{\sigma} & if \;\; S > 0 \\ 0 & if \; S = 0 \\ \frac{s+1}{\sigma} & if \; S < 0 \end{cases} \quad (3.14)$$

The value of the S statistic is associated with the Kendall:

$$\tau = \frac{S}{D} \tag{3.15}$$

Where:

$$D = \left[\frac{1}{2}n(n-1) - \frac{1}{2}\sum_{j=1}^{p} t_j - 1)\right]^{½} \left[\frac{1}{2}n(n-1)\right]^{½+} \tag{3.16}$$

The Z-Transformation equation defined above, for this study, reflects a 95% confidence level, with the null hypothesis that no trend is rejected if Z >1.95. The Kendall tau "τ" is one of the very important Mann-Kendall Statistics as the measure of correlation, which measures the strength of the connection between any two independent variables.

3.4.6 Sequential Mann-Kendall (SQ-MK) Test

The Mann-Kendall test has some restrictions which hinder it from giving the comprehensive structure of a trend for an entire time sequence. There are so many underlying forces throughout a trend and, as there might be fluctuations in the trend over the period under investigation, a sequential test for each and every individual period must be done by applying the specific version of Mann-Kendall test statistics, called Sequential Mann-Kendall (SQ-MK) (Mbatha & Bencherif, 2020).This method has been used successfully by many researchers to detect turning points in trends and their significance (Sneyers, 1991) ; (Loua, et al., 2019); (Mbatha & Bencherif, 2020).
The SQ-MK test technique produces two time sequences, a forward/progressive sequence (u(t)) and a backward/retrograde sequence (u'(t)), and it also has the ability to detect estimated potential turning points in trends in lasting time series, which can be attained by plotting the progressive and retrograde time series in the same graph, if these lines cross each other and diverge beyond the specific threshold (± 1.96 in this study), and if there are statistically significant trends. Mbatha & Bencherif, (2020), computed SQ-MK by rank values, which are y_1 of a given time series ($x_1, x_2, x_{3,..}, x_n$) in the examination, and the magnitude of y_1, (i=1,2,3….,n), which are compared with y_1,

(j=1,2,3,…,j-1). In each comparison, the number of cases where $y_1 > y_j$ is counted and then contributed to n_1. The statistic t_1 is thus calculated by the following:

$$t_1 = \sum_{j=1}^{i} n_i \tag{3.17}$$

The mean and variance of statistic t_1 are given by the following:

$$E(t_i) = \frac{i(i-1)}{4} \tag{3.18}$$

and

$$Var(t_i) = \frac{i(i-1)(2i-5)}{72} \tag{3.19}$$

The forward or progressive sequential values of statistic $u(t_i)$ that are standardised are calculated using the following equation:

$$u(t_i) = \frac{t_i - E(t_i)}{\sqrt{Var(t_i)}} \tag{3.20}$$

To calculate the backward or retrograde statistical values $u'(t_i)$, a similar time sequence ($x_1, x_2, x_3 \ldots\ldots x_n$) is employed, but statistical values need to be calculated by initiating from the end of the time sequence. The forward and backward sequential plotting in one figure make it easy to detect the approximate developing trend commencement. However, it is noted that, sometimes, in the SQ-MK trend evaluation, the forward and backward trends terminate each other, leading to a non-significant trend (Mbatha & Bencherif, 2020)

3.4.7 Multi-Linear Regression (MLR) Model For Checking Connection Between Dependent Variables

To assess further the connection amongst dependent variable, ET, and other parameters, the Multi-Linear Regression (MLR) was used. This model is commonly used to define the connection between a single continuous dependent parameter with other independent variables (Mbatha & Xulu, 2018) . The model has number n output presented as follows:

$$y_1 = \beta_0 + \beta_1 x_2 + \cdots + \beta_p x_{ip} + \epsilon_i \; where \; i = 1, 2, 3 \ldots \ldots .., n \qquad (3.21)$$

The dependent parameter, ET, is represented by (y_i), independent parameters, which are SF, precipitation, NDVI, NDWI and NDDI, are represented by (x_{ip}). β_0 is the intercept, and the coefficients of x are represented by β_1, β_2......β_p. The term ϵ_i stands for the error term that is continuously minimised by the model.

3.4.8 Auto-Regressive Integrated Moving Average (ARIMA) Model

Auto-Regressive Integrated Moving Average (ARIMA) was used in this study. ARIMA is one of the most widely used models because of its statistical properties and it can be used in different ways, including Pure Auto-Regressive (AR), Pure Moving Average, and combined ARIMA series (Kishore & Pushpalatha, 2017). It is also called the Box-Jenkins modelling approach and it is one of the most used time series because of its flexibility, even though it cannot predict non-linear relationships, as its linear correlation structure is presumed among the time values (Zhang, Zhang, & Li, 2016). In a study by Weisang & Awazu, (2008), ARIMA was defined as the model that can be decomposed into two parts, with the first part being an "Integrated (I) component (d)", representing the quantity of distinguishing to be achieved in the sequence to make it constant, and the second being the ARIMA model sequence rendered constant through variation. ARIMA is regarded as the most effective prediction tool and it is used widely in social science and for time series, and it depends on historical data as well as its past error relations for predicting (Adebiyi, Adewumi, & Ayo, 2014). In a study by Gautam & Sinha, (2016), ARIMA was reported as being the most appropriate modelling tool for examining and predicting hydrological events. The model was explained further as the model that explains the linear mixture of the earlier state of a variable "(pure AR component), and previous prediction error (pure MA component)". Therefore, in the current study, the ARIMA model was one of the prediction techniques applied to assist in seeking accurate prediction of evapo-transpiration at the Keiskammahoek Irrigation Scheme. The ARIMA model can be explained mathematically as follows:

$$y = \theta_0 + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \ldots\ldots \emptyset_p y_{t-p}$$
$$+ \varepsilon_t - \emptyset_1 \varepsilon_{t-1} - \emptyset_2 \varepsilon_{t-2} \ldots\ldots \emptyset_q \varepsilon_{t-q} \quad (3.22)$$

Where the terms y_t and ε_t are the actual value and the random error at a given time t. The parameters $\emptyset_i(1,2,\cdots,p)$ $\theta_j(0,1,2,\cdots,q)$ are model parameters. The model parameters p and q are integers and are normally explained as orders of the model. The model random errors, ε_t, are predicted to be independently and identically distributed with a mean of zero and a constant variance of σ^2. The equation above is a general equation that represents several, essential, special cases of the ARIMA family of models. For example, if $q = 0$, then this ARIMA model becomes an AR model of order p. On the other hand, when the parameter $p = 0$, this model reduces to an MA model of order q. Therefore, the most important part of designing the ARIMA model is to determine the appropriate model order (p, q).

3.4.9 Auto-ARIMA

For this study, AUTO-ARIMA models were used to predict evapo-transpiration at Keiskammahoek Irrigation Scheme, using R- Software.

AUTOARIMA using "R" and a portmanteau test called Ljung-Box was done to test the excellence of the time series model. AUTO-ARIMA models do not function as the old, traditional ARIMA models. These models automatically examine numerous permutations of model stipulations and give back the best fitting model. Even though running the AUTO-ARIMA models is similar to running the traditional method, the difference is that the model parameters p, d and q are no longer needed, and their mixture is automatically run and assimilated (Simulator, 2005). According to Burns, (2002), this test is used and, should the significant auto-correlation not be found in model residuals, the model is considered to be perfect. If the values of correlation of residuals for various time lags is not significantly different from zero, the model is then considered to be adequate for use in prediction.

The AUTO-ARIMA function is in the library of the R Programming Software for prediction, and it must define the model parameters, which are p, d, and q, having a minimum AIC condition. The purpose of using AUTO-ARIMA is to conduct an examination of likely models, within the order constraints provided, and it gives back the best ARIMA model according to either AIC, AICc, or BIC value. To measure the relative suitability of AUTO-ARIMA for any statistic model, Akaike Information Criterion (AIC) is used. The AIC values provide the Means for model selection in the examination of the time series. The criterion is: AICek-2km(L), where k is the number of parameters in the model, L is the maximised value of the possible purpose for the predicted model to be use for modelling for that time series. Therefore, the p and q values are selected to minimise Mean AIC, which is (p,q) (Pravilovic & Appice, 2013).

3.4.10 Artificial Neural Network (ANN) Model

ANN is a family of artificial intelligence technique which is capable of prediction and modelling of any time series, including the geophysical time series. ANNs are non-linear data-driven networks that were designed and inspired by the theory of neuroscience (Morimoto, Ouchi, Shimizu, & Baloch, 2007), hence the word “neural” in the name. These are mathematical models based on the capabilities of the human brain to predict and classify problem domains. Khanna, Piyush, & Bhalla, (2014) described ANN as “the information processing paradigm that is inspired by the way biological nervous systems such as a brain process information”.

Artificial Neural Networks (ANN) are respective raw electronic models based on the neural composition of the brain. They are brain modelling which assures fewer alternatives to develop machine solutions. This latest technique to computing also gives greater, needed degradation when there is overload of the system compared with the traditional methods. The basic distilling element of a neural network is a neuron which is made up of construction blocks of human cognisance covering least normal abilities. Generally, a biological neuron accepts inputs from other sources and blends them in another way, performs a generic non-linear assignment on the product and then outputs the final outcome. ANNs can be summarised as techniques that provide a unique method to fix many problems whilst consecutively creating their own needs, (Anderson & McNeill, 1992). The development of ANNs is based on biological neural networks and they have desirable features which include: their adaptivity, heavy parallelism, learning capacity, least energy usage, essential contextual information processing, their deficiency tolerance, generalising capabilities, shared representation and computation. The architecture of ANN is grouped in two categories, explained below, namely feed-forward networks, which do not have loops, and recurrent networks, which have loops because of feedback connections (Jain, Mao, & Mohiuddin, 1996).

- **The feed-forward networks**

The feed-forward networks are also called multi-layer perceptrons which are neurons organised into layers with uni-directional connections amongst them. These networks are static as they produce single sets of results rather than multiple values from an output. Feed-forward networks have very little memory, and they can only rely on the previous network state to respond (Jain, Mao, & Mohiuddin, 1996).

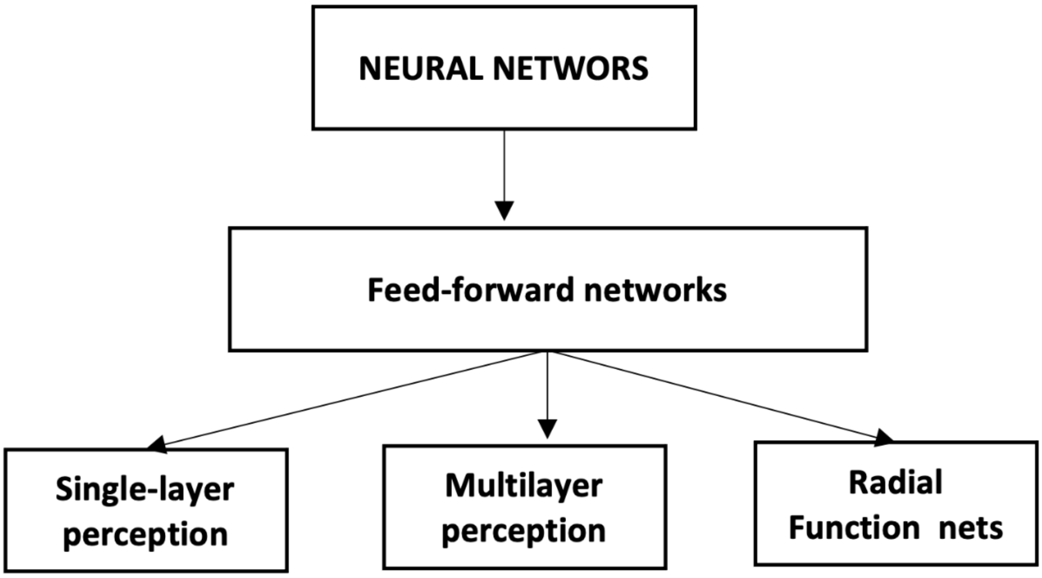


Figure 3.6: Taxonomy of feed-forward networks

[**Source:** (Jain, Mao, & Mohiuddin, 1996)]

- **The recurrent or feedback networks**

These are classified as dynamic systems where, during input presentation, neuron results are computed. This happens because every output is modified because of its feedback paths, which results in a network entering a new state.

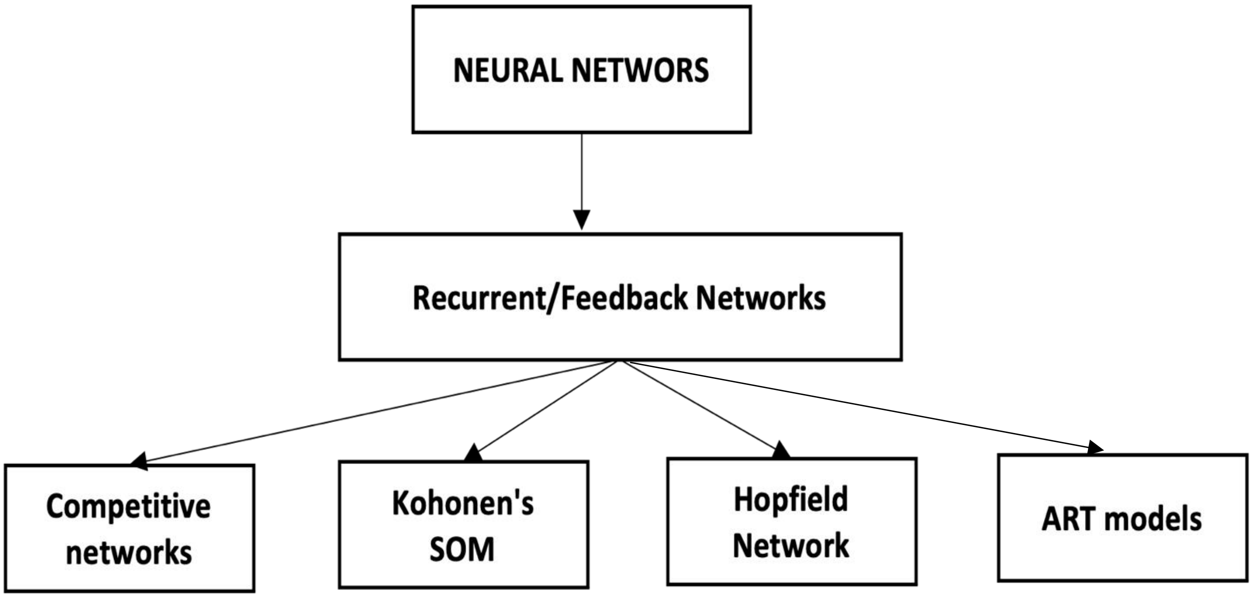


Figure 3.7: Taxonomy of recurrent/ feedback networks
[Source: (Jain, Mao, & Mohiuddin, 1996)]

ANNs are fundamentally a semi-parametric regression method with the capacity to estimate any quantifiable function to an unlimited degree of correctness (Parasuraman, Elshorbagy, & Carey, 2007). They have been adopted widely for predicting and modelling in diverse fields of research such as finance, medicine, engineering, and sciences and also to solve an extraordinary range of problems (Maier & Dandy, 2000). ANNs are specifically useful when the relationships between both input and output variables are discrete Jha, (2007).These models have been commended as being favourable models in cases where the variety of data is excessive and the understanding of the relationship among the variables is mainly unclear (Schultz, Wieland, & Lutze, 2000).

In this study, the single hidden layer feed-forward network was used as one of the techniques to predict ET. Zhang G. P., (2003) explained a single hidden layer feed-forward network as a widely used model for forecasting models for modelling and for predicting time series. The model has three processing parts which are linked by its acyclic layer and distinguished by its connection between output (y_t) and inputs (y_{t-1}.

$Y_{t-2}, .., y_{t-p}$. (Zhang G. P., 2003) gave the following mathematical association between input and output in the model:

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j g\left(\beta_{0j} \sum_{i=1}^{p} \beta_{ij} y_{t-1}\right) + \varepsilon_t, \tag{3.23}$$

Where α_j $(j = 0,1,2,\cdots,q)$ and β_{ij} $(i = 0,1,2,\cdots,p; j = 1,2,\cdots,q)$ are model limits which are called the joining weights; *p* and *q* are the number of input nodes and the number of the hidden nodes, respectively. When designing these types of NNs, the logistic function is often employed as the hidden layer transfer function that is given by:

$$g(x) = \frac{1}{1+\exp(-x)} \tag{3.24}$$

However, it is noted that the ANN model presented above performs a non-linear functional mapping from the past observations $(y_{t-1}, y_{t-2}, \cdots, y_{t-p})$ to the future value y_t given as:

$$y_t = f\left(y_{t-1}, y_{t-2}, \cdots, y_{t-p}, w\right) + \varepsilon_t \tag{3.25}$$

Where w is a vector of all parameters and f is a function determined by the network structure and connection weights (Zhang G. P., 2003).

- **Training the Artificial Neural Networks**

A multi-layer perceptron (MLP) type of network (Figure 3.8) was used, as it is the most used form of neural network. Provided with sufficient data, sufficient hidden units, and sufficient time, an MLP can learn to estimate almost any function to a precise degree (Jha, 2007).

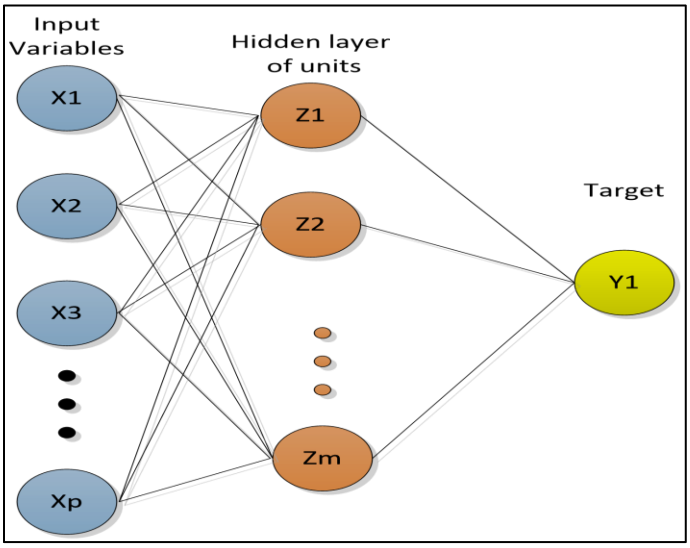


Figure 3.8: Typical structure of ANN

3.4.11 Hybrid (ARIMA-ANN) Model

A Hybrid (ARIMA-ANN) model was used to check the accuracy of the results obtained from two other models that were used, namely, Auto Regressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN). While both models can be satisfactory in modelling and predicting using time series, ARIMA is able to detect linearity of the time series, whilst the ANNs are capable of detecting non-linearity of the time series. Therefore, each model alone cannot adequately detect linear and non-linear patterns but, by using joint models, multi-faceted, auto-correlation structures in data can be modelled precisely (Zhang G. P., 2003). As an example (Mallikarjuna & Rao, 2019) used a Zhang Hybrid model in a study and reported that neither ARIMA nor ANN is completely appropriate for prediction of all the time series because real-world time series have both linear and non-linear correlation structures between observations. Thus, (Mallikarjuna & Rao, 2019) followed a study by (Zhang G. P., 2003) , and used both ARIMA and ANN and developed a hybrid system which is given by:

$$y_t = L_t + N_t, \tag{3.26}$$

Elwasify, (2015) described what each of these values presents as follows:

- Y_t - represents the observation of time series at time t.
- L_t - represents the linear part of ARIMA models, and
- N_t- represents the non-linear part of the ANN models.

According to Zhang G. P., (2003), the first step is to model using ARIMA for the linear component and the residual from the liner data will contain the non-linear relationship. Letting e_t output the residual at time t from the linear model, then e_t is presented as follows:

$$e_t = y_t - L_t, \tag{3.27}$$

Where L_t is the prediction value of time t from the original formula the relationship predicted using ARIMA. This residual is crucial in the diagnosis of the adequate linear models, as the linear model is not adequate should there still be linear correlation structures remaining in the residual. Currently, there is no statistic for non-linear auto-correlation connection diagnosis, which means that, even when models have been accepted by the diagnosis examination, they might still be accurate enough for a non-linear relationship to be modelled properly, and that means every non-linear pattern cannot be modelled by ARIMA. Modelling residual using ARIMA will assist to discover that the relationship is non-linear and (Zhang G. P., 2003) suggested the models for residual as follows:

$$e_t = f(e_t - 1, e_t - 2, \ldots\ldots, e_{t-n}) + \varepsilon_t, \tag{3.28}$$

Where ε_t is the random error whilst *f* is determined by the non-linear function using neural networks and, if *f* is not adequate, that will mean that the error is not certainly random. It is crucial to determine the perfect model and, therefore, by inputting prediction from the residual model, the combined prediction will be the following:

$$\hat{y} = L_t + N_t \tag{3.29}$$

This means that the first step will be to use the ARIMA model to examine the linear part and the second step of in using the Hybrid will be to develop the models using the residual from the first ARIMA model, because the residual from ARIMA will contain the non-linear patterns, and the results obtained from neural networks will be used to estimate the model error in ARIMA terms. Therefore, the Hybrid model, according to (Zhang G. P., 2003), will have different features and the considerable power of ARIMA and ANN and they will determine different patterns.

3.5 COMPARISON OF MODEL PERFORMANCE

Normally there are no standard norms for evaluating the prediction performance of a model and appraisal with other benchmark models (Mbatha & Bencherif, 2020). In this study, evaluating the performance of three models, ARIMA, ANNs, and Hybrid, was done by comparing the predicted ET values with their corresponding ET values, obtained from the site, using typical performance metrics. According to Makridakis, et al., (1982), there have been so many alternative models used over the years to predict the time series equivalent that there are alternative models that can be used. Therefore, it is necessary to choose the appropriate model to use, considering specific conditions. In this study, it was then crucial to check the accuracy of the model to select the most appropriate model based on the predicted ET results. The following performance measures were used: RMSE, MAPE and MAE, as explained by (Karbasi, 2018).

3.5.1 Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(ET^{i}_{estimated} - ET^{i}_{real}).^{2}} \tag{3.30}$$

The RMSE (Root Mean Square Error) is a measure of differences between the predicted results and the obtained results. According to Chai & Draxler, (2014), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) have been used widely in studies to evaluate model assessment. Chai & Draxler, (2014), compared both the RMSE and MAE and their results were contrary to the results for a generic method suggested by (Willmott & Matsuura, 2005), who found that RMSE was not a perfect indicator of average model performance. Chai & Draxler, (2014), found RMSE to be more effective in presenting model performance, which is why this model evaluation was used in the current study to compare the model performances. Chai & Draxler, (2014), also mentioned that not using absolute value was an advantage of RMSE. The error metrics used by this evaluation tool were key because of the tool's ability to distinguish between the model output.

3.5.2 Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{N}\sum_{i=1}^{n}\left|\frac{ET_{estimated}^{i} - ET_{real}^{P}}{ET_{real}^{i}}\right| x100\% \quad (3.31)$$

Mean Absolute Percentage Error (MAPE) was one of the models that was used to measure model accuracy by measuring the quantity of error in predicted values compared with the real values obtained. This measure is used mostly because of its intuitive proportion judgement of error. The measure is used mostly when the amount to predict is known to remain higher than zero (De, Golden, Le, & Rossi, 2016). According to Khair, Fahmi, Hakim, Al, & Rahim, (2017) "the measure is calculated using the absolute error in each period divided by the observed values that are evident for the period, then average those fixed percentages". The approach can be utilised to calculate the borderline error in the forecasted least square method of data analysis and it shows the quantity of error in forecasted values compared with actual values (Khair, Fahmi, Hakim, Al, & Rahim, 2017).

3.5.3 Mean Absolute Error

$$MAE = \frac{1}{n}\sum_{i=1}^{N}\left|ET_{estimated}^{i} - ET_{real}^{i}\right| \quad (3.32)$$

Mean Absolute Error (MAE) was another measure to assess the accuracy of the three selected models in this study. MAE is calculated from an average error, and it is used frequently to evaluate vector-to-vector regression models. In many cases the MAE has out-performed the RMSE in assessing the precision of a model average (Qi, Du, Siniscalchi, Ma, & Lee, 2020). This measure is calculated by summing the absolute values of the error to obtain the sum error and then dividing the total error by a generic

Central University of Technology, Free State

method suggested by (Willmott & Matsuura, 2005), in their study, (Willmott & Matsuura, 2005) identified the advantages of Mean Absolute Error compared with Root Mean Square Error. Willmott & Matsuura, (2005), found that MAE is a more natural evaluator of average error compared with RMSE which is ambiguous. Based on their assessment, (Willmott & Matsuura, 2005) concluded that MAE should be used for any inter-comparison of models' performances. According to (Chai & Draxler, (2014), there are different views regarding whether to use MAE or RMSE to test model accuracy, with (Willmott & Matsuura, 2005), arguing that RMSE is not a perfect measuring tool for model accuracy. Even though their concerns to avoid RMSE were valid, other researchers still believed that avoiding this indicator in favour of MEA was not a solution and suggested that simultaneous use of both measures could have an advantage. Therefore, both of these models were used in the current study, in addition to others, in order to evaluate the model performances more accurately.

3.5.4 Pearson's Correlation Coefficient

Pearson's correlation coefficient was one of the measures used to evaluate the accuracy of the model used for predicting ET at Keiskammahoek Irrigation Scheme. This model will be one of the models Mukaka, (2012), explained the correlation coefficient as the method of statistically evaluating a likely, two-way, linear association amongst two continuous variables. Mukaka, (2012) explained that this model measures the reputed strength of the linear relationship amongst the variables being tested. A correlation coefficient value of zero indicates that there is no linear association between the two variables. A value between +1 and -1 indicates a perfect correlation, the strength of which is indicated anywhere between +1 and -1. The positive value indicates a direct relationship between two variables and the negative value indicates that there is an inverse relationship between two variables (Mukaka, 2012). This index is summarised in the following equation:

$$R = \frac{\sum_{1=1}^{n}(X_i^0 - \overline{X^0})(X_i^p - \overline{X^p})}{\sqrt{\sum_{i=1}^{n}(X_i^0 - \overline{X^0})^2}\sqrt{\sum_{i=1}^{n}(R_i^p - \overline{X^p})^2}} \tag{3.33}$$

CHAPTER 4: RESULTS AND DISCUSSION

4.1 INTRODUCTION

A detailed analysis and interpretation of the results and findings of this study are presented in this chapter. The first step analyses the time series of related parameters in order to identify the trends and variability. The modelling results are obtained from R-Studio programming software.

4.2 TIME SERIES ANALYSIS

Figure 4.1 (a and b) shows the observed time series for 18 years, from 2001 to 2018, for Evapo-transpiration (ET), Precipitation (P), Stream flow (SF), Normalised Difference Vegetation Index (NDVI), Normalised Difference Water Index (NDWI), Normalised Difference Drought Index (NDDI), and the Monthly Volume (MV) of Sandile Dam at Keiskammahoek Irrigation Scheme.

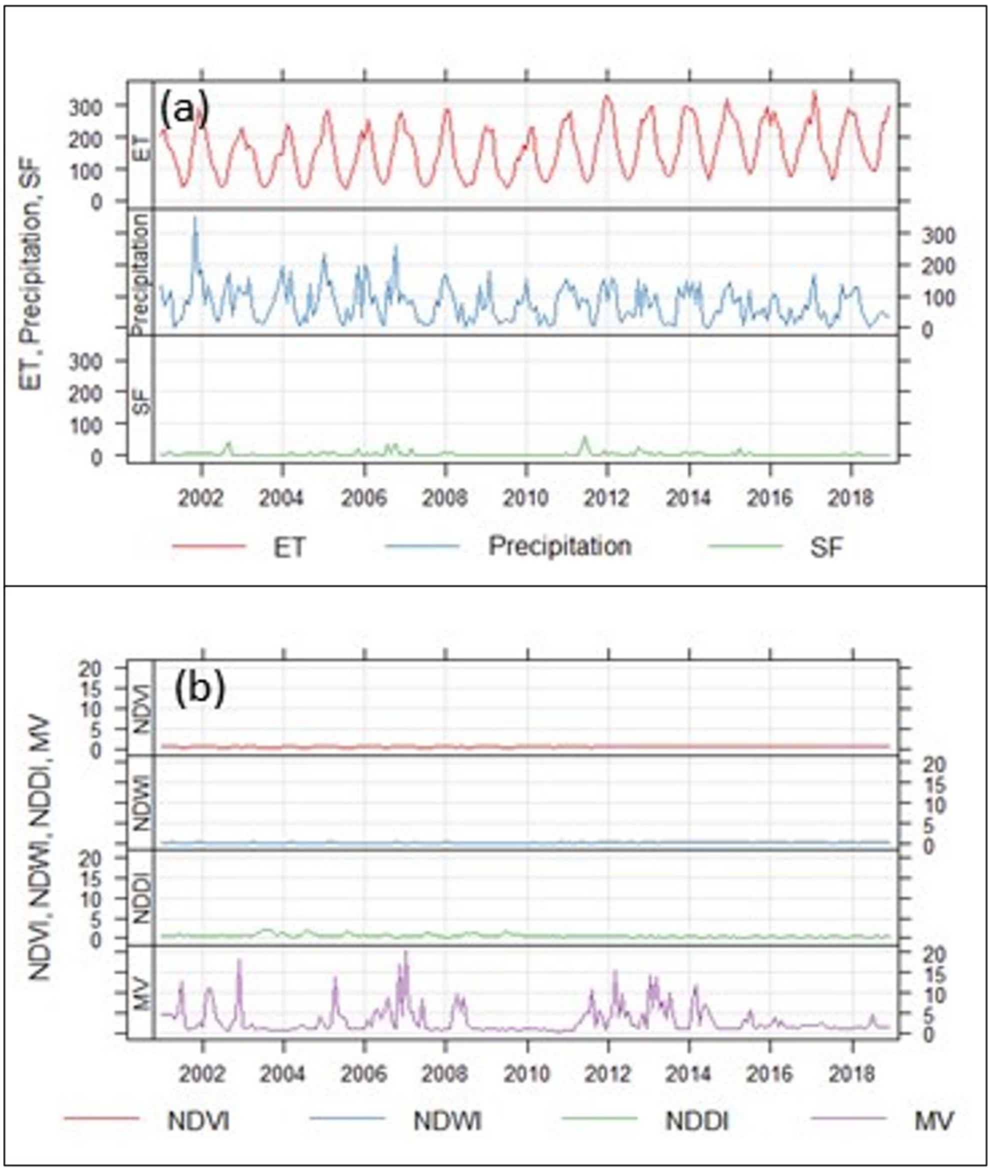
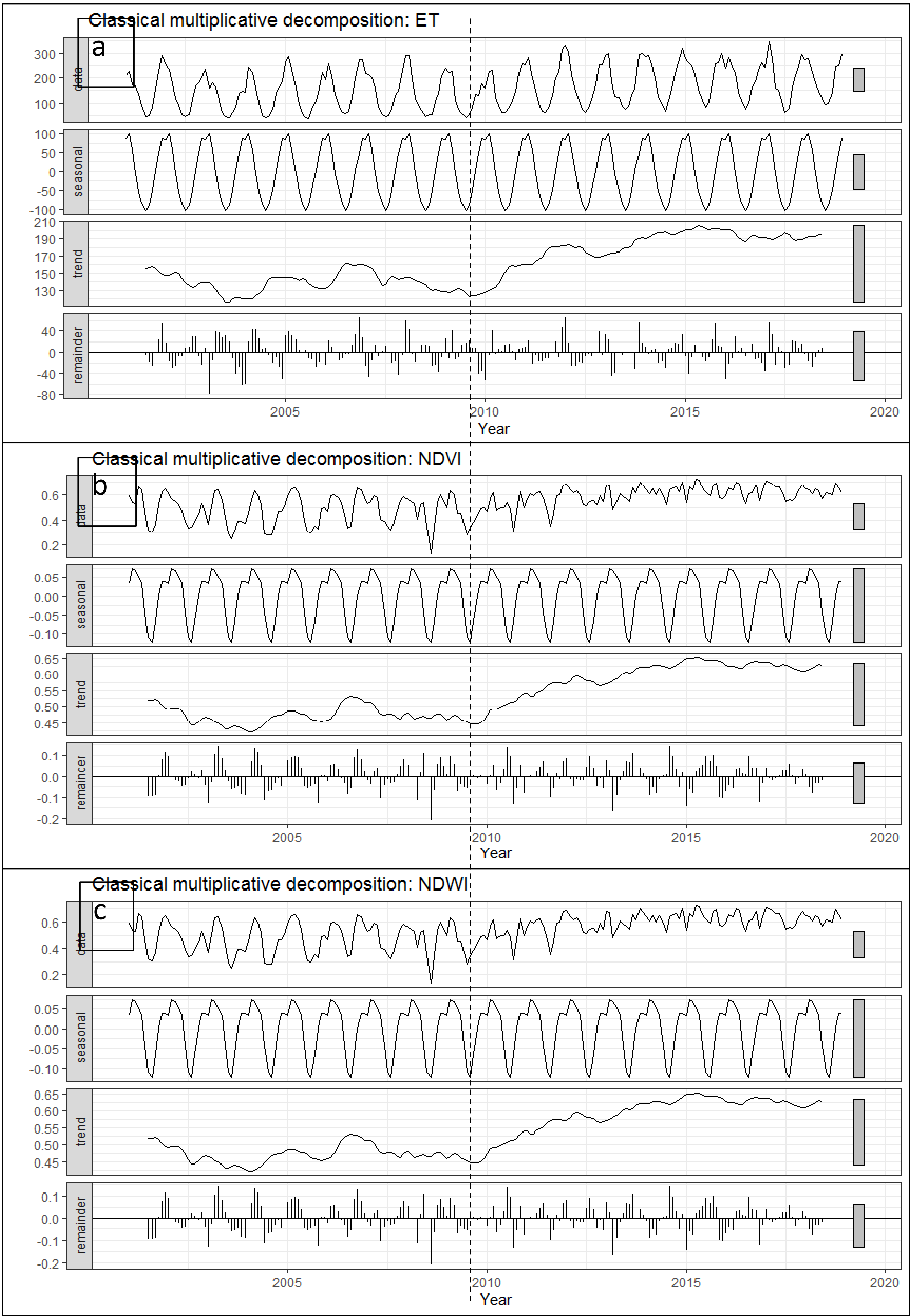


Figure 4.1: Time series for ET, P, SF, NDVI, NDWI, NDDI and MV (2001-2018)

4.2.1 The BFAST Analysis of the Time Series

The time series of all the collected data on the study site was constructed to examine the variability in parameters. Figure 4.2(a-f) shows the temporal variability of ET, NDVI, NDWI, Sandile Dam Monthly Volumes (DM), P and the Keiskammahoek RF volumes. The deconstruction illustrated in Figure 4.2 was performed using the Break for Additive Seasonal and Trend (BFAST) technique.


Classical multiplicative decomposition: ET
a
data
seasonal
trend
remainder
Year
Classical multiplicative decomposition: NDVI
b
data
seasonal
trend
remainder
Year
Classical multiplicative decomposition: NDWI
c
data
seasonal
trend
remainder
Year

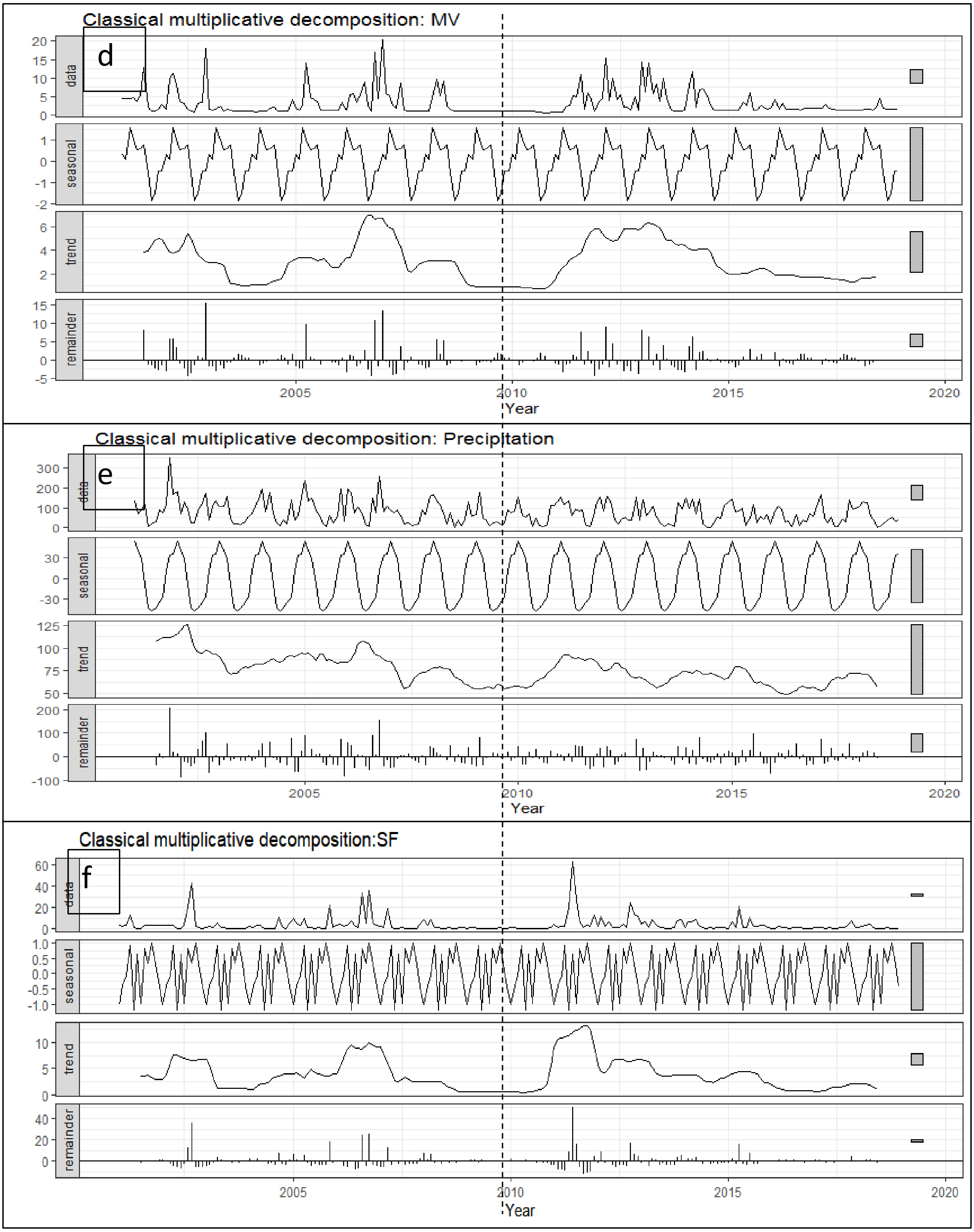


Figure 4.2: BFAST graphs for ET (a), NDWI (b), P (c), NDVI (d), Dam MV (e) and SF (f) time series components

In all Figures 4.2(a-f), the first panel depicts the original time series of the variables followed by the trend on which this analysis is based. It is evident that the annual ET declined during 2001 to 2003, reaching almost zero. During 2003, it is evident that the ET trend in Figure 4.2(a) started increasing in an upward trend that continued till 2007. The trend started declining during 2007 to 2010 but, from 2010 the ET trend started increasing with an upward slope till 2018. This continuous increasing trend indicated that there were irrigation activities happening on the study site.

In Figure 4.2(b), the trend is similar to the ET trend, showing declining green in the study area during 2001 to 2003. During 2003, it is evident that the water content started improving with NDVI increasing until 2007 where the trend starts declining again until 2010. From 2010, the NDVI trend follows the ET trend, which increases in an upward trend until 2018. This full recovery in the greenness of plants over the 18-year period was caused by the irrigation activities which started happening at Keiskammahoek.

Figure 4.2(c) shows the NDWI trend. which is also similar to the ET and NDVI trend over the 18-year evaluation period. This similar change in this variable indicated the extent of water which started to increase in an upward trend from 2010 until 2018.

Temporal variability of the monthly volume of Sandile Dam for the study period was also evaluated. Figure 4.2(d) shows the monthly volume of the Sandile River which is located on the south side of the study site. During 2001, the monthly volume at Sandile Dam shows a decline, with dam levels reaching almost zero. It is evident that, during 2004, the trend started increasing with a spike in 2007 and then the monthly volume decreased during 2009 until 2011, Sandile Dam reaching the lowest levels which started to improve.

Precipitation was also one of the variables that was considered for evaluation. Figure 4.2(e) shows the precipitation trend decreasing continuously during 2001 to 2010. This figure shows unstable precipitation during 2010 to 2018 and this is evident in the declining dam volumes during these years.

The Keiskammahoek River, which is the feeder river for Sandile Dam, was also evaluated. Figure 4.2(f) shows that the trend declines during 2001 until 2003. From

2003 this figure shows that the stream flow volume started improving with an increasing trend until 2007. However, from 2001 the stream flow volume started decreasing again with the lowest stream flow volumes being at Keiskammahoek River during 2007 to 2011. During 2011, the stream volume started increasing, reaching the highest volume in 2011. However, the monthly volumes of river stream flow showed a constant decrease until 2018.

4.2.2 Wavelength Analysis

Figure 4.3(a and b) represent the normalised wavelet power spectra for the monthly mean ET and Precipitation for the period from 2001 to 2018. The “u”-shaped solid line in the figure indicates the cone on influence (COI), defining the area of spectrum that must be deliberated in examination. The COI represents the region in which edge effects transpire. The black lines indicate the 95% confidence level which is significant for that region. The blue colour presents the lowest wavelet power and the red colour represents the regions which have more wavelet power. The vertical axis shows the period (in months) and the horizontal axis represents the time in years (Loua, et al., 2019).

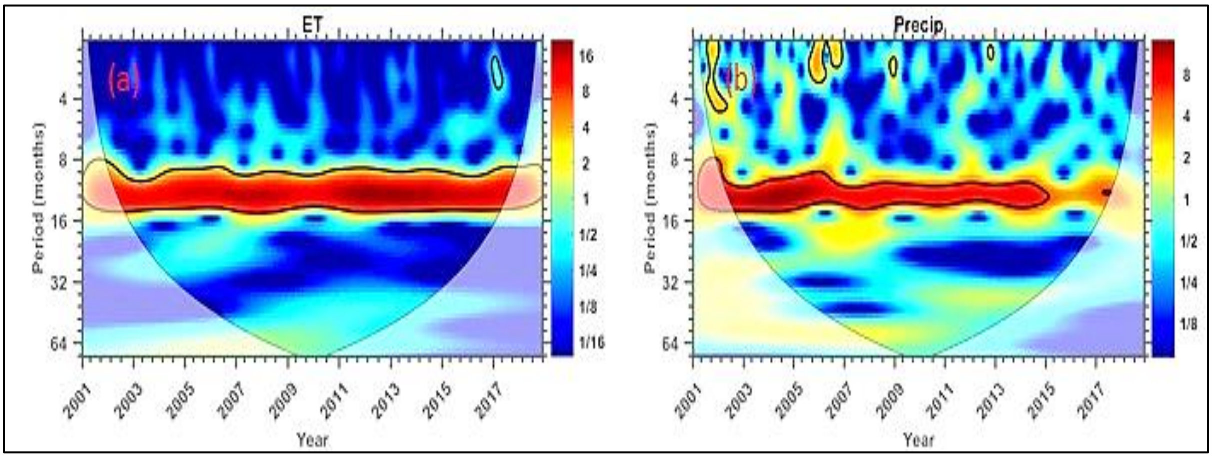


Figure 4.3: The normalised wavelet power spectra of yearly mean ET (a) and Precipitation (b) variability during 2001 to 2018 at Keiskammahoek Irrigation Scheme

Figure 4.3(a) shows the strong wavelet power spectra peak in the 12-month period from 2001 to 2018. However, in the wavelet transforms of the Precipitation, it is evident that the 12 months’ power spectrum was strong during the period 2001 to 2015 and wavelet power spectra were very weak during 2015 to 2018. This weak seasonality power spectrum might have been because of the decrease in Precipitation that occurred, with Precipitation reaching almost 0 mm in 2015. This reduction in precipitation is presumed to be because of one of the strongest *El Nino* events that occurred during the period from late 2014 to 2016 (Mbatha & Xulu, 2018). Precipitation power spectra also experience several shorter periods of significant power primarily during 2002 and 2006.

4.2.3 Wavelet Coherence

Figure 4.4 below shows the cross-wavelet power spectra for ET and Precipitation with phase connection indicated by arrows. The area where two cross wavelet factors are in phase is indicated by arrows point to the right, and the anti-phase if the arrows point to the left, and the ET lagging or leading if arrows point downwards or upwards correspondingly (Loua, et al., 2019). The black solid line indicates the cone of influence (COI), the edge effects develop significant frequency scales, and the line also indicates the significant areas of 95% confidence level.

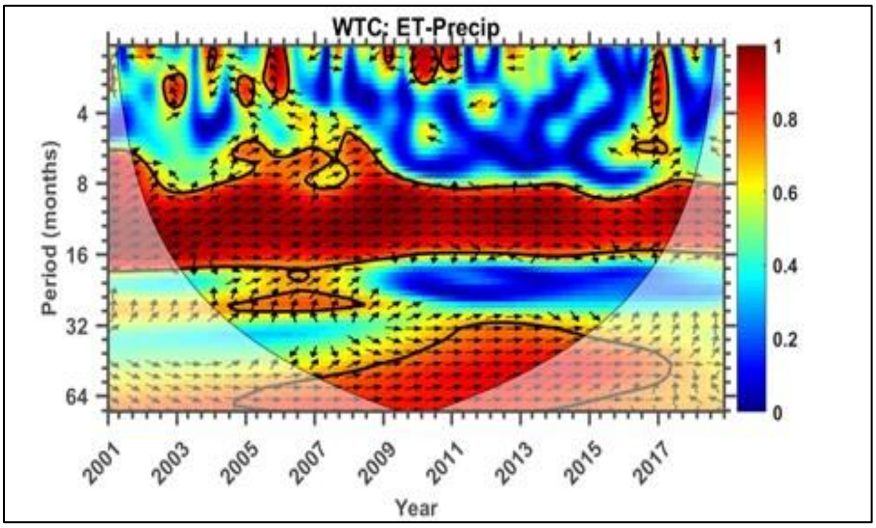


Figure 4.4: Wavelet coherence between ET and Precipitation for period (2001 to 2018) with phase relationship depicted by the arrows.

In general, it is confirmed that there is a possible strong and significant, in-phase relationship between ET and Precipitation during 2001 to 2018, indicated in the period bands 8 months to 16 months, with arrows pointing right (phase). It is evident also in band 1 to 5, there is an anti-phase relationship between ET and Precipitation, and this happened from years 2003 to 2006 and from 2009 to 2010. However, this changes in year 2017 where ET leads precipitation with arrows pointing up. The upward arrows, then rotating to the right in the time period bands 24 – 32 months show that ET leads Precipitation during the period 2001 to 2009. However, in the period bands 34 to 64 months, with arrows pointing downward, showing lagging, and arrows turning right, showing a phase relationship from 2001, reaching 95% confidence level during 2007 to 2015, indicates a very strong phase relationship with Precipitation leading ET. During the period from 2015 to 2018, it is evident that there are some arrows pointing right, showing a phase relationship between the two variables but, there, ET seems to be leading with arrows pointing up.

4.2.4 Correlation Statistics

To assess the relationship between ET and other parameters used in this study, the Pearson Correlation Coefficient was used. The heat map shown in table 4.1; encapsulates the linear connection between all the variables used in this study.

Table 4.1: Pearson correlation coefficient for NDDI, ET, NDVI, NDWI, SF and Precipitation (P).

Parameters collected	NDDI	ET	NDVI	NDWI	SF	P
P	-0.16	0.53	0.23	0.12	0.37	1
SF	-0.13	0.02	0.06	0.09	1	
NDWI	-0.87	0.58	0.88	1		
NDVI	-0.76	0.67	1			
ET	-0.51	1				
NDDI	1					

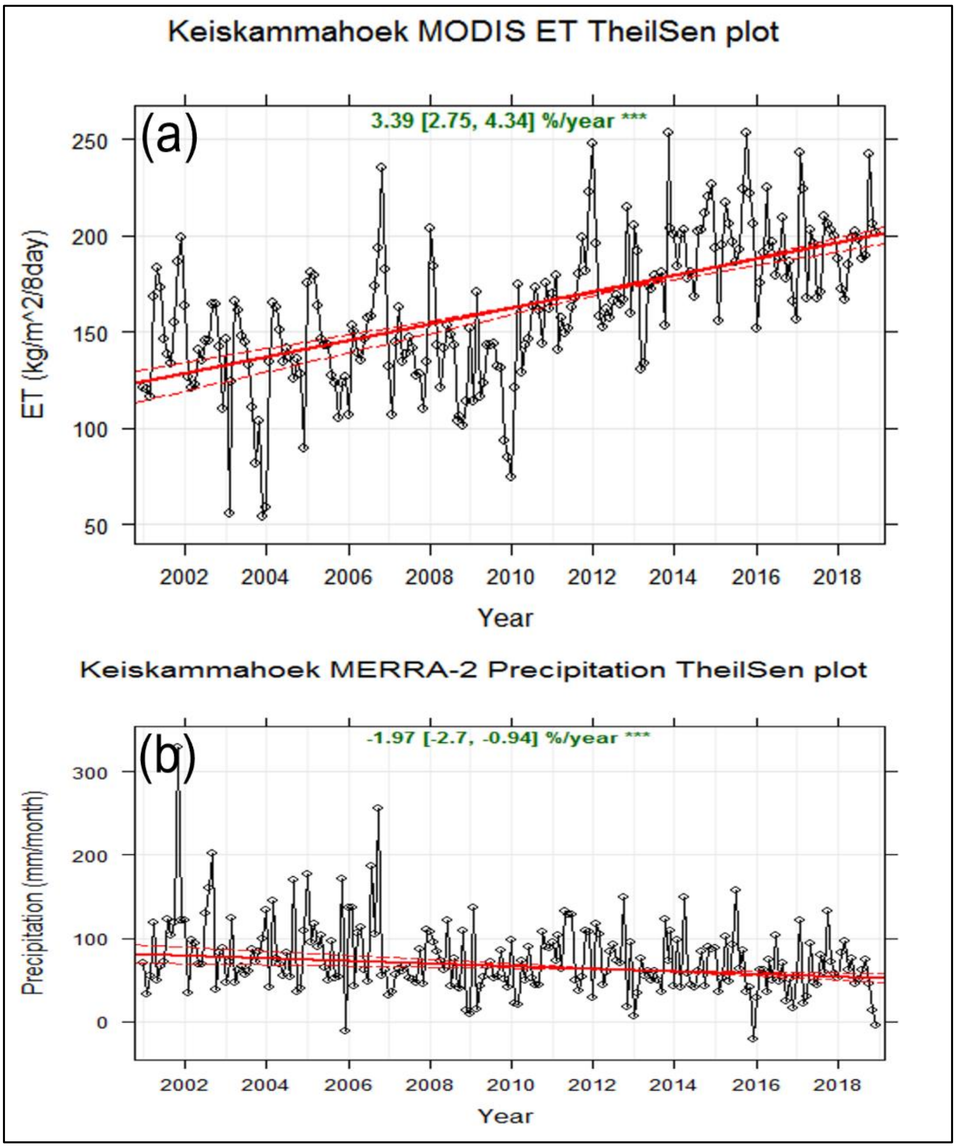
Table 4.1, shows evident that there is a strong negative correlation between ET and NDDI (-0.51), signified by the ET correlation relationship with SF (r = 0.02). However, in the above table, it is also confirmed that ET has a strong relationship with other parameters tested with NDVI (r = 0.67), NDWI (r= 0.58), and Precipitation (r = 0.53). This negative correlation between NDDI which is the drought index simple justify that there was no severe drought on the study site with P(-0.16), SF (-0.130), NDWI (-0.87), NDVI (-0.76) and ET (-.51). According to Lee, et al., (2016), this index is best

indicator of drought as include vegitation and water content, henceforth should be assessed in dry seasons/ dry surface in order to obtain strong correlations. Keiskammahoek Iiirigation Scheme is under irrigation activities which justifies these weak correlations between NDDI and other indices. This is justified by the strong correlation between ET and NDWI (0.58), NDVI (0.67), which indicate that vegitation was not under streesed on the study area, which contradict with results found using the theil-Sein estomator finding significant decrease in Precipitation with $p < 0.001$.
The results of correlation between NDVI and Precipitation (0.23) are intresting as they agree with Palacios-Oruetes, Khanna, Litago, whiting, & Ustin, (2006), where NDVI did not closely follow Preecipitation results, in particular on irrigated regions. In this study even thogh there a correlation between ndvi and Precipitation but it is not strong, which is contary with results found by (Mingjun, et al., 2007), where they found similar correlation between NDVI and Precipitation. Other studies like (Tavazohi & Nadoushan, 2018), proved that decrease in water reserved like Precipitation showed a decrease in correlation between NDWI and other indexes. (Aksoy, Gorucu, & Sertel, 2019), mentioned that, drought is caused by precipitation deficit and increased evapotranspiration and based on the time series data observed, the decrease in precipitation and streamflow raises a concern.

4.2.5 Theil-Sen Plot

Figure 4.5(a-d) shows the time evolution of the de-seasonalised monthly mean for SF, MV, ET and Precipitation time series, with the trend cautioning line superimposed on it. The Theil-Sen function in the study used 216 trend points. The solid line shows the estimate and the dashed red line depicts 95% confidence intervals of the trend based on the variables. The complete trend shows: for Stream Flow (a) -2.15 m^3/year at confidence levels from -3.4 to 0.27 m^3/year; -0.14 m^3/year for the Monthly Volumes of Sandile Dam (b) at confidence level from -1.37 to 1.19 m^3/year; for ET (c) 3.39, at confidence level from 2.75 to 4.34 km/m^2/year (a); and -1.97 mm/year at 95% confidence level from -2.7 to -0.94 mm/year for Precipitation (d).

The value of p determines the symbols used and that is represented by $p < 0.001$ = ***, $p < 0.01$ = **, $p < 0.05$ = * and $p < 0.1$ = +. The three stars indicate a highly statistically significant trend estimate; but the plus sign signifies a trend prediction that is not statistically significant (Bencherif, et al., 2020). In Figure 4.5, the dashed red lines designate the 95% confidence level of the trend.



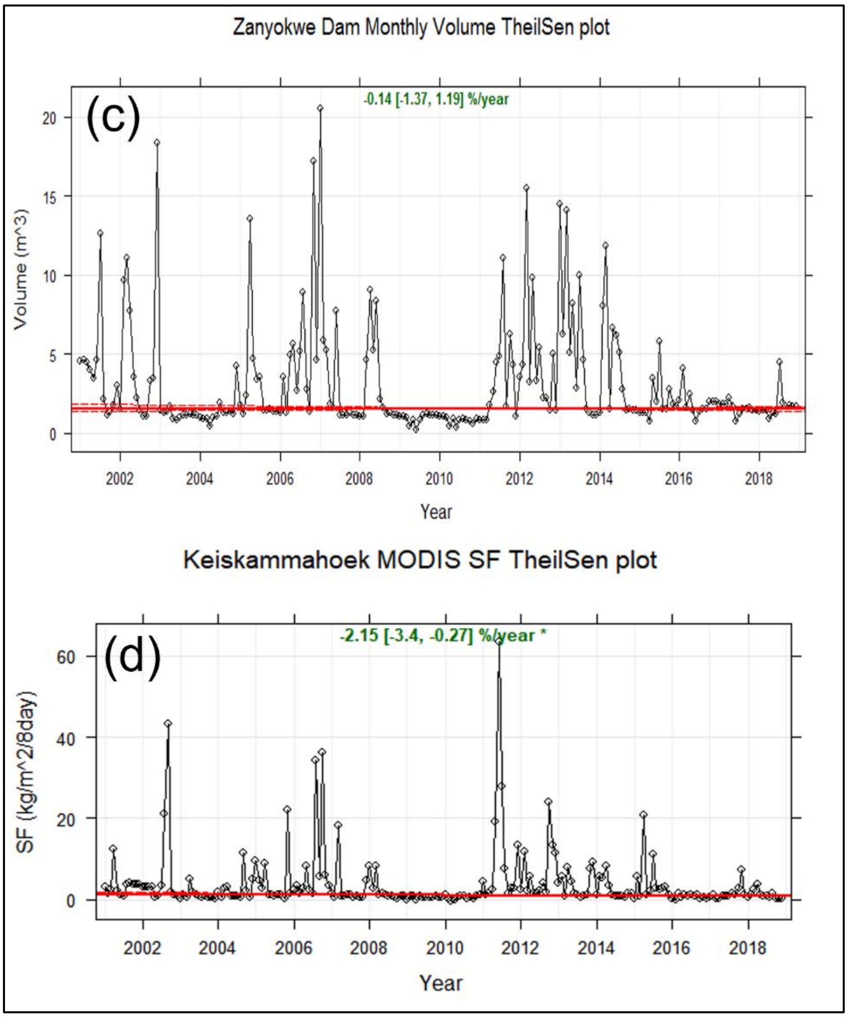


Figure 4.5: Long-term trend of monthly SF, Sandile Dam MV, ET (a) and (P), MV (c), SF (d).

The Theil-Sen trend estimation for the time series of the SF in Figure 4.5(a) indicates a significant downwards trend of the stream flow, with a p-value of approximately -2.15 m^3 /s). Figure 4.5(b) shows the evaluation of the Sandile Dam level, which is the lower dam feed by the Cata Dam currently used for irrigation in the Keiskammahoek Irrigation Scheme. The results presented in this figure indicate a downwards trend of approximately -0.14 m^3/year. In terms of the significance of this trend estimation, and according to the Theil-Sen function, it indicates a less significant downwards trend of the dam's levels, with a p-value of less than 0.1. This river is important for this study because it feeds water to the Sandile Dam that supplies water for irrigation in the study area. In this figure, the annual evolution of ET displays a positive slope, which corresponds to the increasing trend (see Figure 4-5(c) at 3.39 kg/m^2 per year, and this upsurge is strongly significant with $p < 0.001$. Figure 4.5(d) shows a downward slope for the Precipitation trend for the period tested, confirmed by the negative value of total precipitation per year (-1.97 mm/year). This downwards trend indicates that there is a significant decrease in precipitation with $p < 0.001$.

4.2.6 Man-Kendall Test

The Mann-Kendall (MK) non-parametric test method was used in this study in order to investigate the ET and Precipitation trend further and its significance for the data taken from the Keiskammahoek Irrigation Scheme. Table 4.2 shows the Mann-Kendall test model for ET and Precipitation with z-score results for the period of 18 years from 2001 to 2018. According to (Xulu, Peerbhay, Gebreslaie, & Ismail, 2018), the Mann-Kendall trend test model is used to quantify the trend's significance, with a z-score between -1.96 and +1.96 indicating non-significance of the trend and, if the variable falls outside of the range, the trend is regarded as being significant. For this study, this trend test was employed with ± 1.95 boundary lines which specify 95% confidence levels.

Table 4.2: Long-term trend of monthly SF, Sandile Dam MV, ET (a) and (P), MV (c), SF (d).

Variables	Z-score	Kendal's tau	S	Var(S)	P-value
ET	3.898	1.782946e-01	4.140000e+03	1.127460e+06	9.698e-05
Precipitation	-2.6134	-1.195521e-01	-2.776000e	1.127460e+06	0.008964
SF	-1.7508	-8.016553e-2	-1.860000e+03	1.127368e+06	0.07997
MV	**0.60747**	**2.785445e-02**	**6.460000e+02**	**1.127391e+06**	**0.5435**
NDVI	8.3291	3.809626e-01	8.845000e+03	1.127455e+06	2.2e-16
NDWI	**10.021**	**4.583118e-1**	**1.061200e+04**	**1.127460e+06**	**2.2e-16**
NDDI	-9.8859	-4.521102e-01	-1.049800e+04	1.127460e+06	2.2e-16

For ET, it is shown that the z-score value is equal to +3.898, a value far greater than +1.96, and this indicates a significant positive trend with a cut-off line of 95% confidence level. This indicates that, over the years, ET has been increasing in the study area. By contrast, the z-score for precipitation is equal to -2.6134, indicating a significant decrease in precipitation during the study period of 18 years. This suggests a strong decrease in precipitation variability at Keiskammahoek Irrigation Scheme over this period. The SF z-score for Keiskammahoek River is equal to -1.7508, which signifies a non-significant decrease in streamflow for this river over the 18-year period. Corresponding to that, the z-score of the Sandile Dam is equal to +0.60747, also indicating a non-significant increase in the dam level over the study period. It is evident also that there was a significant increase in NDVI over the 18-year period with a z-score reaching +8.3291, which corresponds with a strong significant increase in NDWI over this period, with a z-score equal to +10.021. The last variable which was tested using the Mann-Kendall trend test model was NDDI. NDDI has proven to be a useful index to evaluate drought, because it uses information on both vegetation and water conditions (Gouveia, Bastos, Trigo, & DaCamara, 2012). It is evident that there was a significant decrease in NDDI during the 18 years period with a z-score equal to -9.8859 for Keiskammahoek Irrigation Scheme.

4.2.7 Sequential Mann-Kendall (SQ-MK) Test

While the Mann-Kendall Test is useful for indicating the trend of a total time series and its significance, it is beneficial to know the dynamics of the trend's significance and the change detection points of a trend. This was done using a specialised Mann-Kendall test called the Sequential Mann-Kendall (SQ-NK). Figure 4-6(a) shows the sequential statistics value of forward/progressive (Prog) *u(t)* (red solid line) and backwards/retrograde (Retr) *u(t)* (solid black line), calculated using SQ-NK for the ET time series extracted from Keiskammahoek Irrigation Scheme. In general, there is an upward trend evident in the SQ-MK parameters from the beginning of the time series.

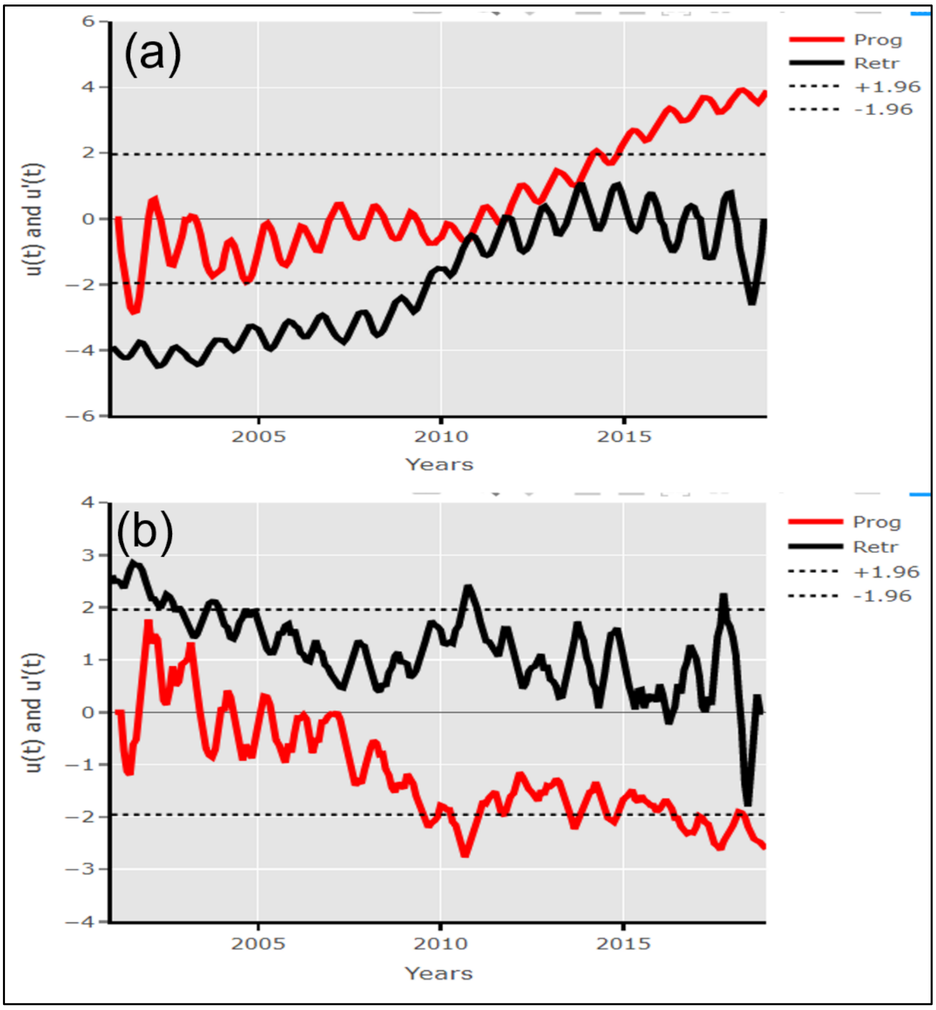


Figure 4.6: Sequential Mann-Kendall (SQ-MK) test for ET and P data for the period from 2001 to 2018.

In Figure 4.6(a), the change detection point is evident during the period between 2012 and 2014. After the change detection points during this period, the progressive time series increases until it reaches a significant level during 2015. The time series of the progressive statistic reaches z-score values which are approximately equal to 4 towards the end of the time series. This indicates that there has been an increase of the evapo-transpiration rate on the study site that started just after year 2010. This increase is presumed to be associated with a continual increase in average surface temperatures as well as the intense *El Nino* events that affected the eastern part of South Africa in recent years (Mbatha & Xulu, 2018).

Figure 4.6(b) depicts the sequential statistic values of forward/progressive (Prog) *u*(t) (solid line) and retrograde (Retr) *u*'(t) (black solid line) found by using the Sequential

Mann-Kendall (SQ-MK) test for Keiskammahoek Irrigation Scheme Precipitation data for the period from 2001 to 2018. During the period from 2001 to 2010, it is evident that there is a decrease in trend that reaches a significant level immediately after year 2010 in terms of the progressive statistic. After this period, the (Prog) values fluctuate between above -1.96 and below -1.96, which generally indicates several incidences of a negative and significant trend.

Ripin et al. (1989) suggested in a study that any variation in climate change affects both ET and Precipitation, and any decrease in Precipitation would lead to an increase in Evapo-transpiration. In the current study, the SQ-MK statistics indicate a significant decrease in precipitation with a downward trend indicated by (Prog) during the period 2016 to 2018. From 2001 to September 2002, it is evident that Precipitation decreased with a significant downward with (Retr) deceasing from 2.613 to 2.099. (Prog) also indicated a non-significant decrease in precipitation from the year 2001 to May 2010, but this decrease became significant and reached a maximum value of (-2.77) in September 2010. Both (Retr) and (Prog) start decreasing until 2018, even though some insignificant trends are evident in both (Retr) and (Prog) values. It can be concluded that the SQ-MK test for ET data shows ET at Keiskammahoek Irrigation Scheme is subject to an increasing trend with no significant decrease, and the effect of irrigation activities on the study site and led to the increase in ET becoming a significant trend. Parallel to that Precipitation is subject to decrease even though, most of the time, both (Retr) and (Prog) values are not significant. The last results indicate a significant decrease with negative (Prog) -2.613.

4.2.8 Multi Linear Regression Analysis

To elucidate the connection between ET and other variables, the Multiple-Linear Regression analysis (MLR) was used in this study. This model is famous for its ability to explain the connection between a single continuous dependent parameter and two or more supplementary independent parameters (Mbatha & Xulu, 2018).

Table 4.3: Multi-Linear Regression (MLR) model where ET is the dependent variable and SF, P (P), NDVI and NDWI are independent variables.

Parameters	Estimate	Std. Error	t value	Pr- Value	Sig
ET	3.4076197	0.2009740	16.956	< 2e-16	***
SF	-0.0156514	0.0032221	-4.857	2.32e-06	***
Precipitation	0.0045486	0.0004318	10.534	< 2e-16	***
NDVI	2.7564601	0.4022261	6.853	7.89e-11	***
NDWI	-0.3202777	0.5270840	-0.608	0.5441	
NDDI	-0.2901542	0.1246038	-2.329	0.0208	*

Table 4.3 shows the MLR analysis statistics encompassing ET, SF, Precipitation, NDVI, NDWI and NDDI. Other hydrological variables that have an influence on ET were evaluated to determine their degree of influence on ET at the study site. It is evident that there is a significant relationship between ET and SF, Precipitation and NDVI with their p-values of 2.32×10^{-06} for SF, 2.32×10^{-06} for Precipitation, and 2×10^{-16} for NDVI. This correlation demonstrates a statistically significant relationship between ET and the three variables since their ρ-values are less than 0.05. However, the statistical relationship between ET and NDDI was found not to be very significant with p-value of 0.0208, which is greater than 0.05, but it is still significant. It is also indicated on Table 4.2 that the relationship between ET and NDWI is statistically insignificant with p value of 0.5441, which is greater than 0.05.

4.3 MODEL SELECTION AND DISCUSSION OF RESULTS

For the purpose of this study, only the ET data were considered for prediction and modelling because one of the main objectives of the study is to predict evapo-transpiration in the study area. Three model types were considered, namely, ARIMA, ANN and the Hybrid ARIMA-NNAR. The results obtained by using these models are discussed in this subsection. The model flowchart in Figure 4.7 below depicts the process that is followed in this section.

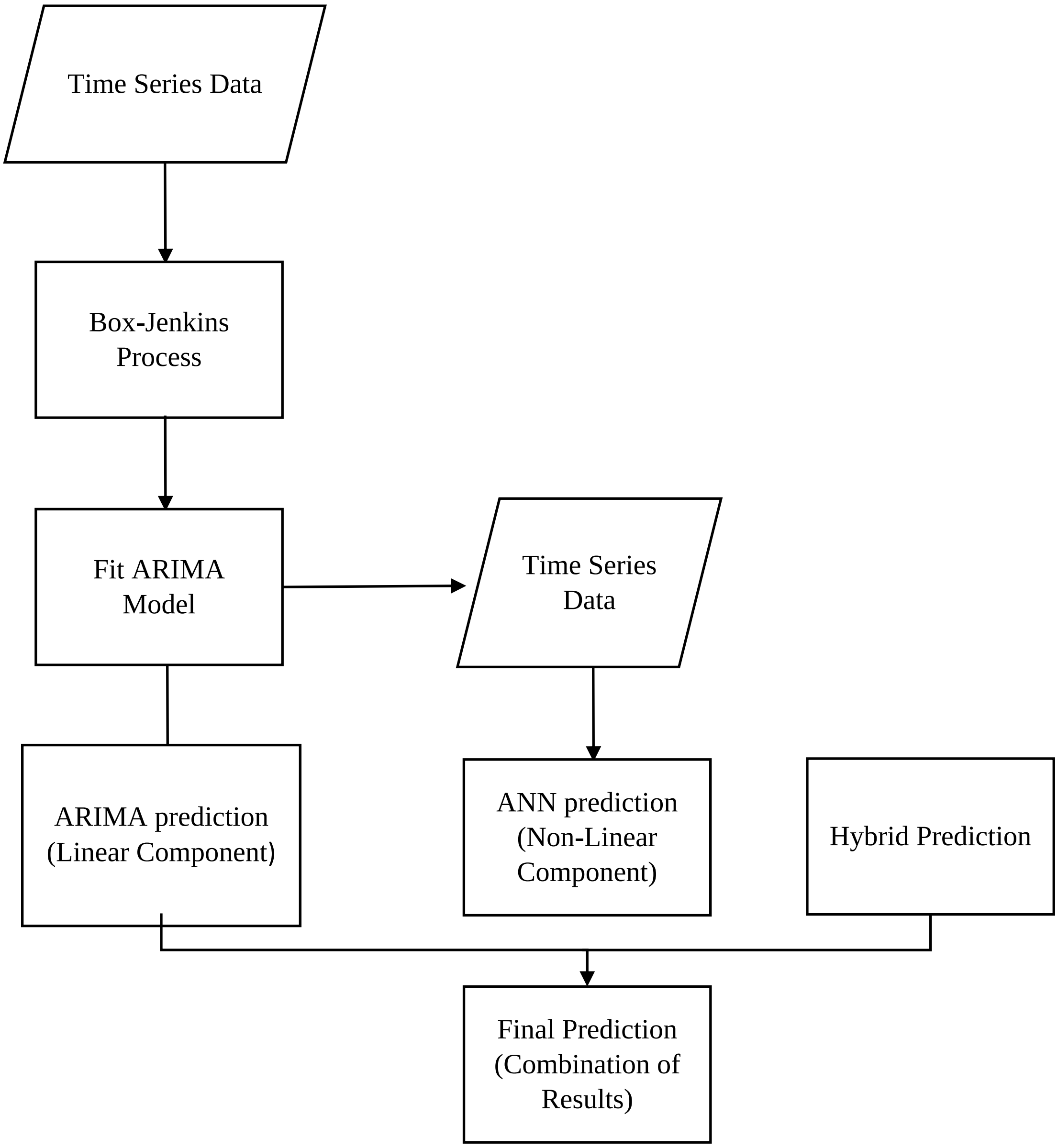


Figure 4.7: Model flow chart

4.3.1 ARIMA Model Training and Validation

Figure 4.8 shows the Akaike's Information Criteria (AIC) graph that indicates that there is no significant correlation because all the bars do not exceed the dotted line indicating the 95% confidence level and, according to (Widowati, Putro, Koshio, & Oktaferdian, 2016), and (Gautam & Sinha, 2016) the residue is random. The best ARIMA model selected to predict ET is ARIMA (1,0,0). Figure 4.8 shows residuals which are evenly dispersed. The normal distribution of residuals indicates that the selected ARIMA model is free of overfitting (Reza & Debnath).

In this study, the training of the ARIMA model was done by selecting data from 2001 to 2015 as a training set of data. One of the important aims of splitting the data into a training part and a testing part is to use the testing part of the time series to check the sign of the variable's parameters, and also to investigate whether they are significant or not.

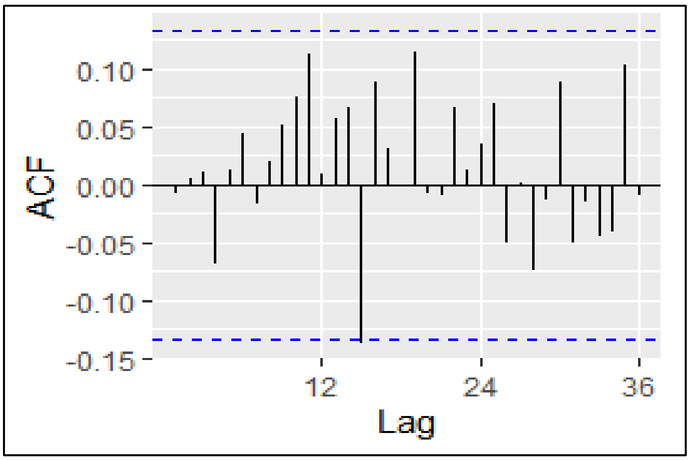


Figure 4.8: Auto-Correlation Function (ACF) of residual as an ideal fitted model for data series of ET from 2001 to 2018.

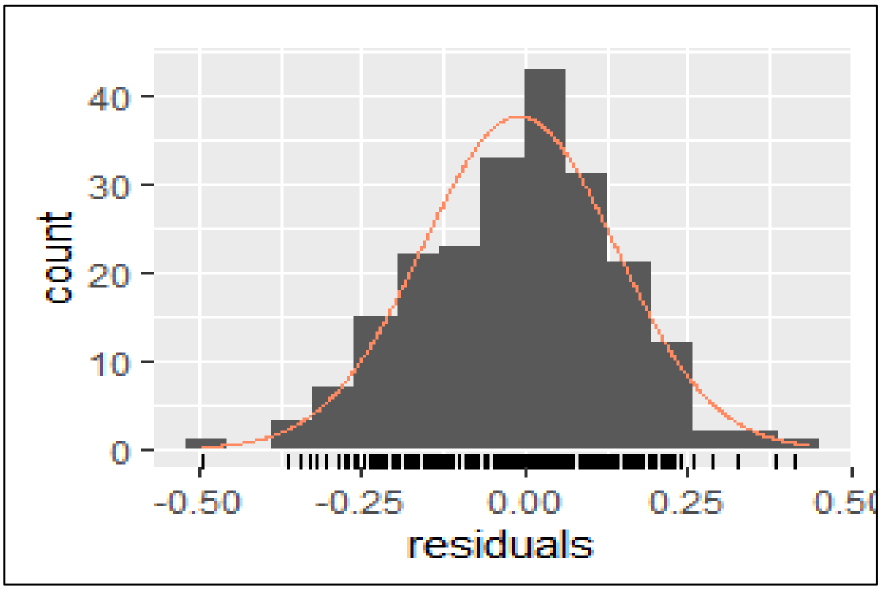


Figure 4.9: Histogram of residuals as a best fitted model for data series of ET from 2001 to 2018.

Figure 4.10 (a-b) shows the training part of ARIMA model. This figure shows the raw data fitted to the ARIMA model. The black line indicates the ET time series from (2001 to 2018), with prediction indicated by the red line also from (2001 to 2018). The same figure also indicates the prediction done for 5 years from (2019 to 2023). This was to test whether the model could be able to predict and the model successfully proved that it could predict the future before the actual prediction take place. The training for the hybrid (ARIMA-ANN) was not done as this index is a combination of ARIMA and ANN.

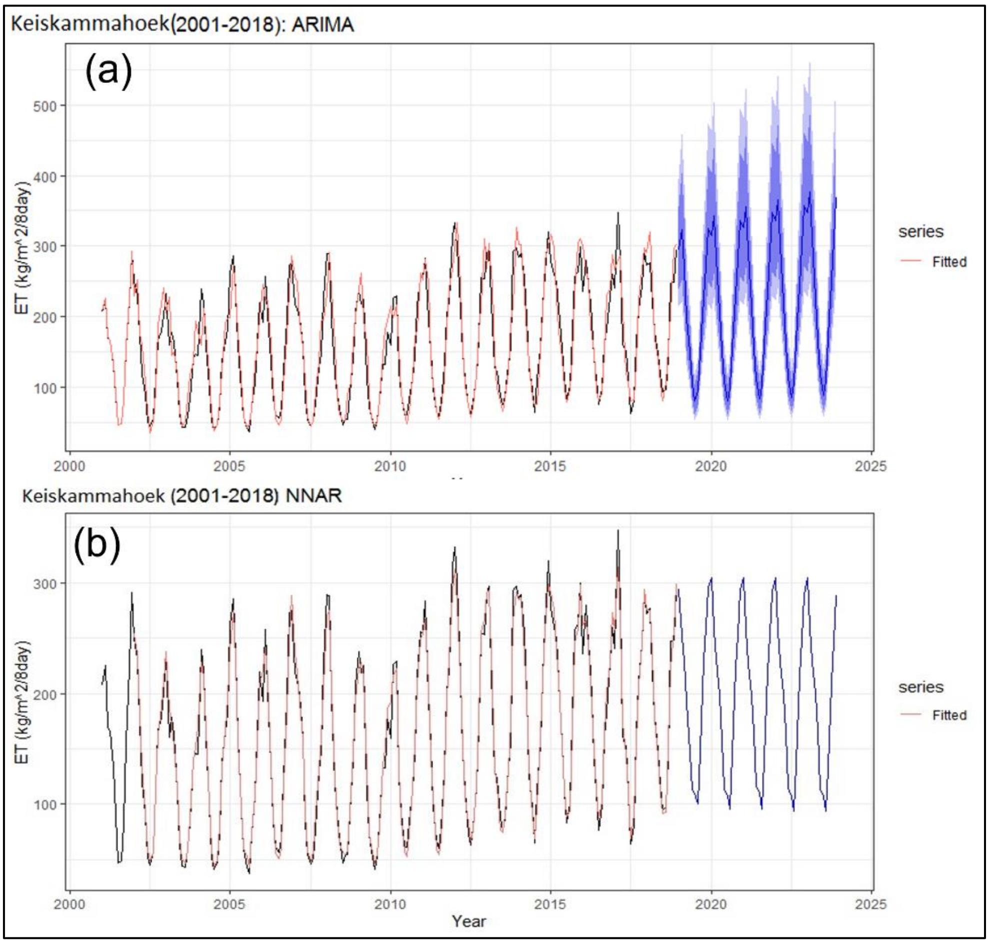


Figure 4.10: ARIMA and ANN training using data from (2001-2018)

The model accuracy was then tested in order to assess the accuracy of the actual data (fitted) and the forecast in order to see if the results will follow the same trend or otherwise. Based on following results, it is evident that based on this models test, ANNs proves to be the effective model with less values of RMSE (10.61), MAE (8.09), MAPE (5.68) and with higher correlation between the fitted and the forecasted results.

Table 4.4: Comparison of ARIMA and ANN: RMSE, MAE, MAPE and R

Models	RMSE	MAE	MAPE	R
ARIMA	25.41	18.80	12.13	95
ANN	10.61	8.09	5.69	99

4.3.2 ARIMA Prediction

After the training of the model, using 15 years of data from 2001 to 2015, the next step was prediction using the remaining 3 years of data from 2016 to 2018. Thus, the data set from 2016 to 2018 was used as the testing part of the time series. This is important in prediction because the testing part is predicted and then prediction results are compared with the truth.

Figure 4.11 shows the training part of the ET time series, indicated by the black line (2001 to 2015), and the ET predicted results from 2016 to 2018, indicated by the blue line, with the dark grey and light-grey shadings indicating the 80% and 95% confidence levels of the predicted time series. The ARIMA model constructed for this data is ARIMA (1,0,0). To assess correlation between the predicted results and the actual ET data, the correlation coefficient between the two datasets was calculated. This correlation coefficient was found to be R = 0.94, which indicates a strong correlation between the predicted ET data and the testing time series.

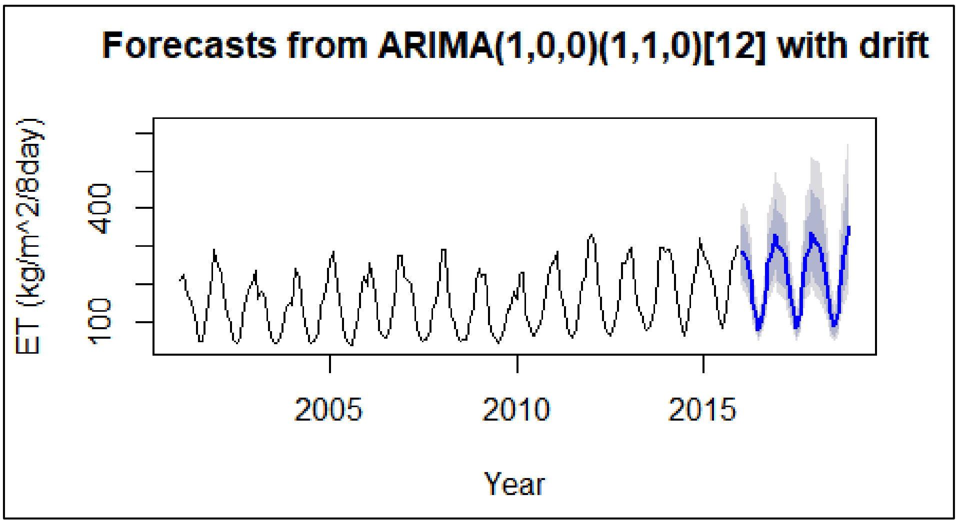


Figure 4.11: Predicted evapo-transpiration for 3-year period from 2016 to 2018 using the ARIMA model.

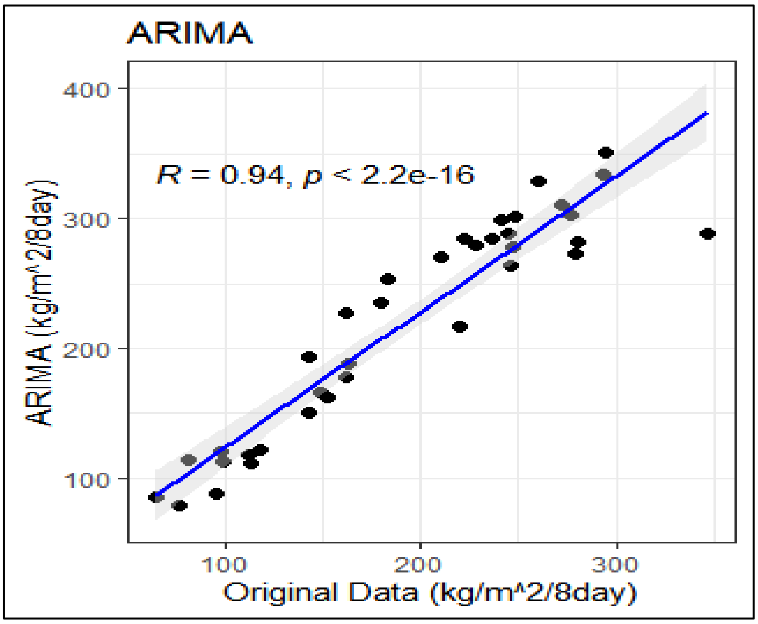


Figure 4.12: Scatter plot between observed and predicted values using ARIMA model (with validation period of 2016 – 2018).

Based on the good correlation between the predicted results and the observed results shown in Figure 4.12, it is evident that the ARIMA models prove to be adequate tools to predict ET at Keiskammahoek Irrigation Scheme. The good correlation results shown in Figure 4-12 are encouraging and they are consistent with the results of (Gharde, Kothari, & Mahale, 2016), where ARIMA models proved to be able to predict with R more than 0.9 and achieving less values of RMSE. It is noted further that the lower RMSE value (37.58) shown in Table 4-5 below agrees with the study of Mittal, 2011), where RMSE was used to assess the feasibility of ARIMA models in predicting ET through the use of RMSE. These results are very inspiring and prove ARIMA to be an effective tool for predicting linear time series similar to the findings of (Valipour, 2012) ,where ARIMA models also proved to be suitable tools for predicting meteorological data. What is more interesting in this study is the accuracy of ARIMA models in predicting ET compared with ANN models. These results are similar to the results of (Kishore & Pushpalatha, 2017), where ARIMA models out-performed ANNs.

4.3.3 ANN Models Training and Prediction

The “prediction” package of R statistics was used to train the ANN model, using the data for 15 years from 2001 to 2015, which is 83.33% of the observed data.

The NNAR (1,1,2) model was used to predict the remaining data for two years from 2016 to 2018. Figure 4.13 shows the blue prediction portion of the plot for the period of two years (2016-2018). To check the correlation of the prediction portion, (Adler & Ingela, 2010) was used with predicted ET variables against the observed ET variables. Figure 4.14 shows predicted ET and observed ET represented by black dots falling on the diagonal through the origin. The Pearson Correlation confirms the strong relationship between these variables with R = 0.86.

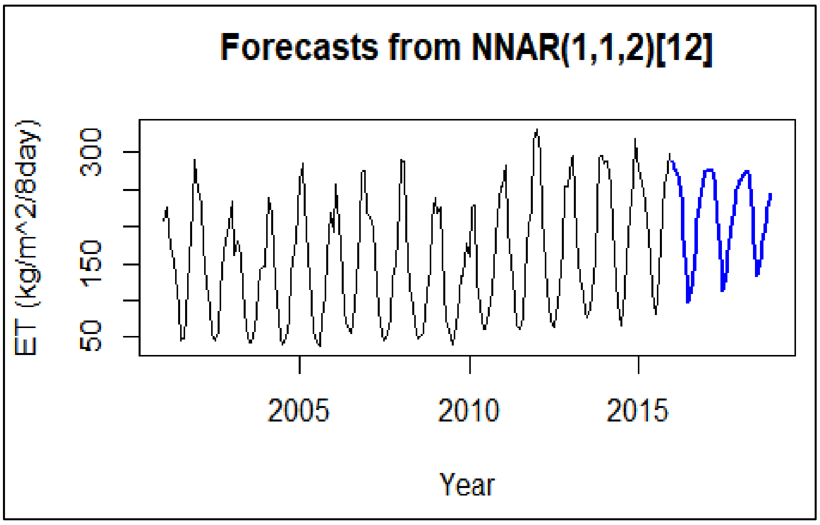


Figure 4.13: The actual data from 2001 to 2015 and the predicted evapotranspiration results using NNAR models from 2015 to 2018

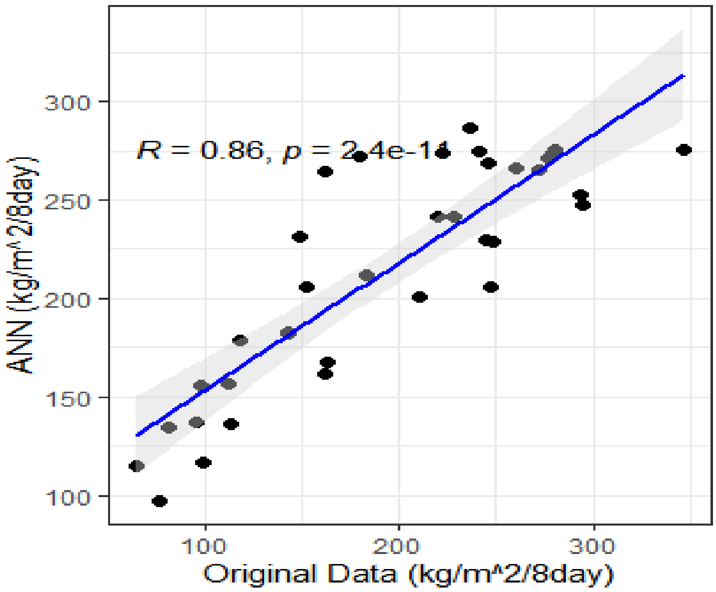


Figure 4.14: Scatter plot between observed and predicted ET using NNAR model (with validation period of 2016 to 2018)

In this study, ANNs were employed as the second prediction model to assess the non-linear components of the ET time series. The scatter plot ,shown in Figure 4.14 above, demonstrates the very good correlation between the ANN observed results and predicted results for ET at Keiskammahoek Irrigation Scheme. The effectiveness of ANNs prediction of ET is similar to the study by (Ogunrinde, Oguntunde, Fasinmirin, & Akinwumiju, 2020), where the viability of ANNs for predicting ET was proved. These results differ somewhat from the study of (Kihoro, Otieno, & Wafula, 2004), where the ability of ANN and ARIMA was compared in predicting a monthly time series using various datasets for international airline passengers, tourists and people visiting Japan and Kenya, not data for mean monthly temperatures for air. This empirical comparative assessment proved that ANN performed better than ARIMA models, but (Kishore & Pushpalatha, 2017) stated that the nature of the data might affect the results. Based on correlation comparison between ARIMA and ANN results, it is proven that ARIMA models out-perform ANN models.

4.3.4 Hybrid (ARIMA-ANN) Model Prediction

The third model that was used in this study is the Hybrid model, which is built by the combination of ARIMA and ANN models. This was done to accommodate both the linearity and non-linearity characteristics of the time series. This model proposed in a study by Zhang G. P., (2003) and it was shown that this model system has the ability to predict both linear and non-linear, underlying processes. The Keiskammahoek Irrigation Scheme also experiences real-world time series, which contain both linear and non-linear correlation structures. For the purpose of applying this model system, the data were divided into a training set comprising 80% of the data (2001-2015) and testing set comprising 20% of the data (2016 - 2018). Figure 4.15 shows the ET data from 2001 to 2015 (training dataset) indicated by the black line, and the blue line which indicates the predicted ET data with the grey shading indicating the 95% confidence levels for the two-year period (2016 to 2018). It is indicated Figure 4.15 that the Hybrid (ARIMA-ANN) model can capture the seasonality in the data, and the visual inspection of the predicted section of the time series indicates the good performance of the model. Thus, to assess the performance of the model, a linear relationship between the predicted section and the training data set was assessed through the estimation of the correlation coefficient “R”. Figure 4.16 depicts the scatter plot that shows the correlation between the predicted ET values and observed data of the ET testing part. In Figure 4.16, it is evident that the correlation coefficient shows a strong linear relationship with R = 0.94. This indicated the excellent performance of the Hybrid (ARIMA-ANN) model system in predicting and modelling the ET time series.

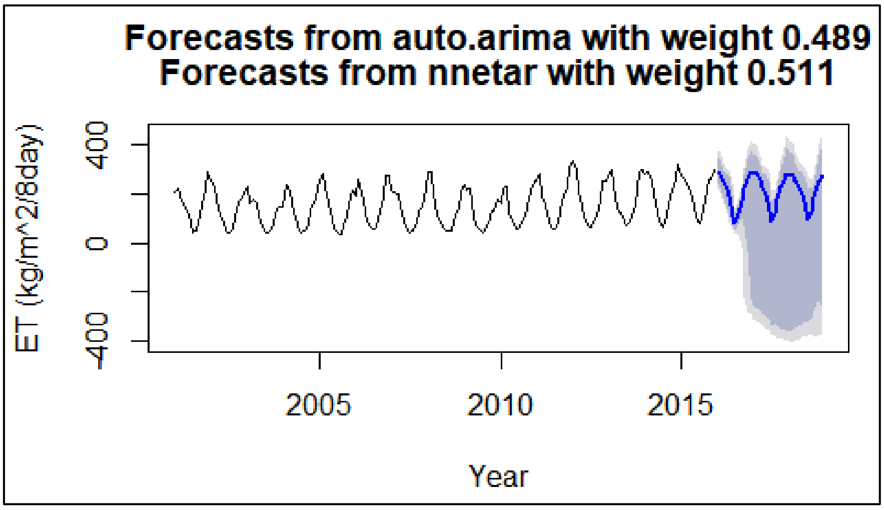


Figure 4.15: Observed ET versus predicted ET from 2001 to 2015 using the Hybrid (ARIMA-NNAR) model (black line indicates observed data and blue line indicates predicted data)

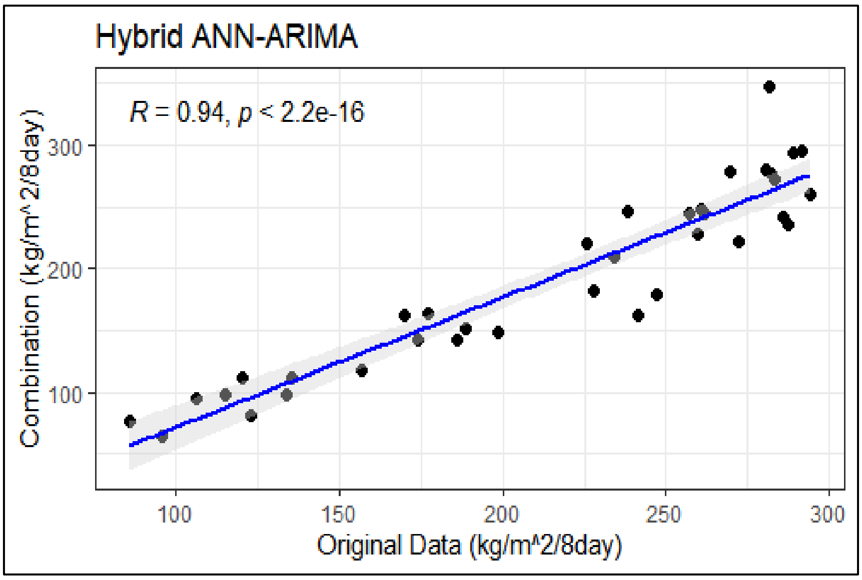


Figure 4.16: Scatter plot between observed and predicted ET using Hybrid (ARIMA-ANN's) model (with validation period of 2016 to 2018)

The Hybrid (ANN-ARIMA) model result shows high correlation between the observed results and the predicted results at Keiskammahoek Irrigation scheme with R = 0.94, which is more accurate compared with other models. The results of this study, where Hybrid (ARIMA-ANN) shows more accurate results compared with other models, are similar to the results of the studies by (Aslanargun, Mammadov, Yazici, & Yolacan, 2007), (Ayub & Jafri, 2020), where this combined model was used to predict different variables and more good prediction results were archived compared with using individual ARIMA and ANN models. Similarly, in a study by Lu, Ntu and Jila (2004), where ARIMA models, ANN models and Hybrid (ARIMA-ANN) models were used to predict technology in predicting the electric short-term load. The Hybrid (ARIMA-ANN) once more proved to be better in performance accuracy. In Table 4.5 below, the excellent performance of the Hybrid (ARIMA-ANN) models is confirmed with lower values of RMSE at 33.80, MAE at 27.02 and MAPE at 17.31. It is very interesting that as proposed by (Zhang G. P., 2003), the results of this study also confirmed that the Hybrid (ARIMA-ANN) model out-perform other prediction models.

4.3.5 Averaged (ARIMA, ANN and Hybrid (ARIMA-ANN) Model Prediction

In this study, three different modelling systems were used, and their performances were demonstrated in terms of the correlation coefficient “R”. However, according to Bates & Clive, (1969) and Clemen, (1989) it is important to average the prediction in order to improve its accuracy. Therefore, in this study, the predictions of the three model systems (ARIMA, ANN and Hybrid ARIMA-ANN) were averaged further to represent the combined model output. Figure 4-17 depicts the combined model results for ET for the time period from 2016 to 2018. Similar to the other models, the time series indicate a strong seasonality in the ET data. The correlation coefficient was also estimated by finding the linear relationship between the combined model and the original testing part of the time series. Figure 4.18 shows the scatter plot for the combined model vs the original data. It is shown in Figure 4-18 that there is a strong relationship between the combined models and the original time series which is indicated by a correlation coefficient of R = 0.94.

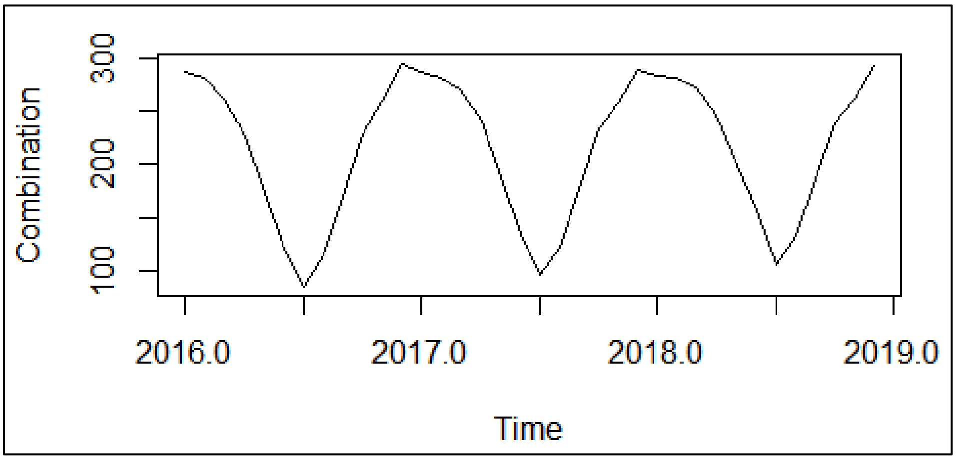


Figure 4.17: Shows predicted ET data from 2016 to 2018. The black line indicates the prediction using the averaged models (ARIMA, NNAR and Hybrid) models

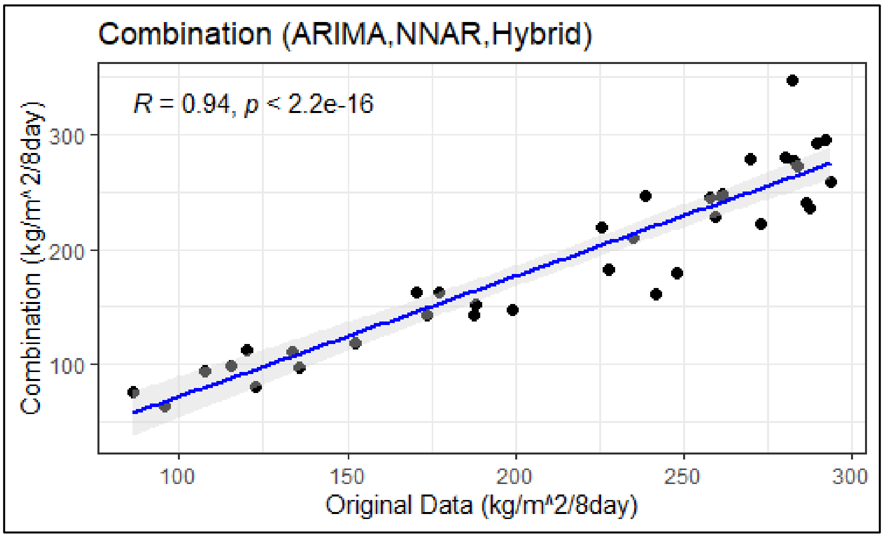


Figure 4.18: Scatter plot depicting observed ET and the predicted ET using the averaged models.

The results of the averaged models, which are promising, agree with the results of averaged predictions done by (Shang & Booth, 2020), who used averaged models in order to improve the accuracy of prediction. Shang and Booth (2020) used empirical model weights that optimise the accuracy for individual prediction horizons, using data from approximately 17 countries. Shang & Booth, (2020), found that averaged models show a promising future, using the known models, and indicated that further investigation is needed. The results in the current study indicate averaged models to the second most promising model that can be used to predict ET for Keiskammahoek Irrigation Scheme which is very interesting.

4.3.6 Comparison of Models

After prediction using ARIMA, ANN, Hybrid, and the combined averaged models, evaluation of the performance of the models was carried out by using model performance equations explained in Chapter 3. A study by Lewis (1982), showed that there are several alternative statistical methods to evaluate model performance, which can be selected according to specific circumstances.

Table 4.5 shows three models used in this study to prediction ET at Keiskammahoek Irrigation Scheme, namely, ARIMA, ANNs and Hybrid (ARIMA-ANN), and the average of the three models. The model prediction capabilities are compared by using model performance statistics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Correlation Coefficient (R). The results presented in Table 4-5 indicated that the Hybrid model outperforms other models with RMSE = 33.80, MAE = 27.02, MAPE = 17.31 and R = 0.93. It is also noted that the Mean Absolute Percentage Error values for ARIMA and Hybrid appear to be similar to ARIMA (MAPE = 17.26) and Hybrid (MAPE = 17.31). Since the Hybrid model is made up of a combination of ARIMA and ANNs, it is possible that this model will perform better than the other models because it is expected to be capable of capturing both the linearity and non-linearity in the time series. In terms of the correlation coefficient, ARIMA seems to outperform the other models, with the correlation coefficient of R = 0.94.

Table 4.5: Comparison of the ARIMA, ANN, Hybrid and Combined Models: RMSE, MAE, MPE, and R.

Models	RMSE	MAE	MAPE	R
ARIMA	37.58	32.37	17.26	0.94
ANN	44.18	35.88	24.35	0.86
HYBRID	33.80	27.02	17.31	0.94
COMBINED	34.68	28.00	18.15	0.94

It has been shown in previous studies that a, combination of multiple prediction methods leads to increased accuracy of the prediction (Clemen, 1989). Therefore, in this study, the predictions obtained by the three models used were combined by using a summation method. The results of the Combined models, shown in Table 4-5, indicate better results than ARIMA and NNAR. These observations are encouraging as they are consistent with what was presented by (Hyndman & Athanasopoulos, 2018) in their study of the combination of several time series prediction methods. Similar to what was obtained in the current study, (Hyndman & Athanasopoulos, 2018) pointed out that the combined predictions often lead to results that are close to, or better than, the best component method.

4.4 PREDICTING FUTURE ET USING THE HYBRID (ARIMA-ANN) MODELS

The main purpose of this study was to predict evapo-transpiration for the study area in order to assist the farm managers to plan efficiently. The application of three prediction models, Auto-Regressive Integrated Moving Average (ARIMA), Artificial Neural Networks (ANN) and the Hybrid model consisting of Auto-Regressive Integrated Moving Average and an Artificial Neural Network (Hybrid RIMA-ANN), assisted in selecting the best performing model based on the widely used statistical model evaluation methods: Root Mean Square Mean Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Pearson Correlation Coefficient (R). The purpose of using different model performance methods was to ensure that model selection is not biased, based only on one statistical measure but based on overall performance from all measures.

The best model selected, which is the Hybrid (ARIMA-ANN) model, was then applied in predicting the future ET for five years from 2018 to 2023. Figure 4.19 depicts the ET observed data from 2001 to 2018, indicated by the black line, and the blue line which indicates the predicted ET for 5 years, with the grey shading indicating the 95% confidence level for the 5-year period (2018 to 2023).

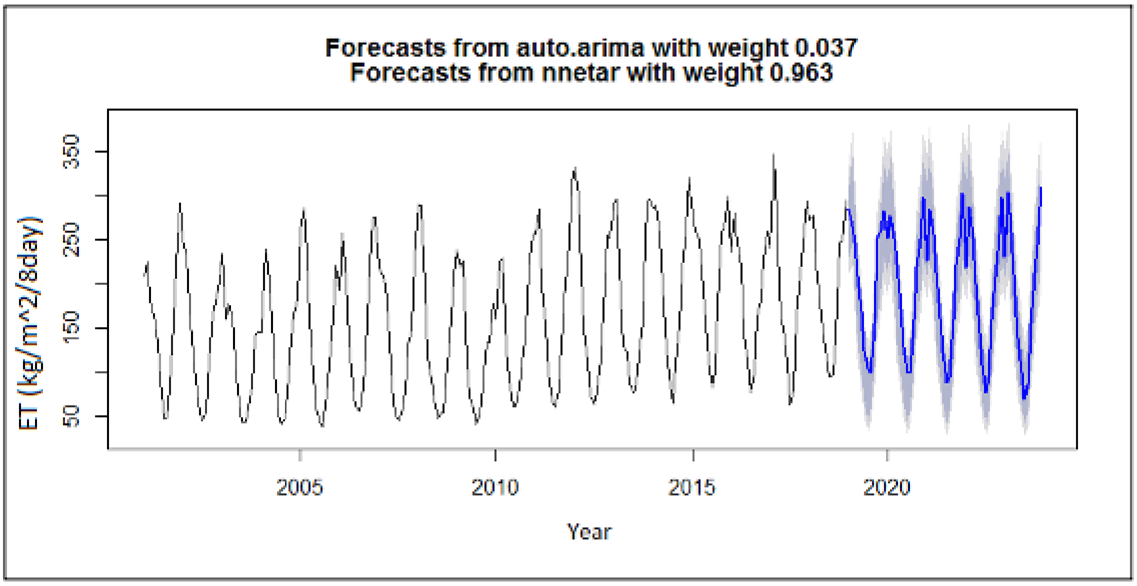


Figure 4.19: Observed ET from 2001 to 2018 and predicted ET for five years (2018 to 2023) using the Hybrid (ARIMA-ANN) model.

It is very exciting to see the Hybrid (ARIMA-ANN) model out-performing other models. This proves that the Hybrid (ARIMA-ANN) model predicts ET with higher accuracy and proves that, in the absence of data used by the traditional Penman-Monteith approach, the Hybrid (ARIMA-ANN) model could be used as a better alternative to predict ET where linear and non-linear ET time series are involved.

CHAPTER 5: CONCLUSION AND RECOMMENDATIONS

5.1 SUMMARY

The possibility to predict evapo-transpiration (ET) is essential as it can affirm the optimum planning, design and operation of any irrigation scheme. Thus, the main aim of this study was to predict evapo-transpiration (ET) at Keiskammahoek Irrigation Scheme, using three time-series prediction models: Auto-Regressive Integrated Moving Average (ARIMA), Artificial Neural Networks (ANN) and the Hybrid (ARIMA-ANN) using data for 18 years from 2001 – 2018. There were three major objectives for this study: to collect and analyse the obtained time series for all the variables; to develop ARIMA, ANN and Hybrid (ARIMA-ANN), to predict ET using these three models; and to assess the performance of these models in predicting ET at Keiskammahoek Irrigation Scheme. The data for 18 years (2001 – 2018) for ET and other related variables, including Precipitation, Normalised Difference Vegetation Index (NDVI), Normalised Difference Water Index (NDWI), were extracted from cloud-built software called Moderate Resolution Imaging Spectroradiometer (MODIS) Tera/Aqua 16-day dataset. The Keiskamma River Stream Flow (SF) and monthly volumes of Sandile Dam were obtained from the Department of Water and Sanitation (DWS) website and Stream Flow was collected from Station R1H015-RIV.

Before the prediction of the ET time-series data, a detailed analysis of all the time-series data used in this study was performed, using various time series analysis methods in order to understand the behaviour of the time series. A tele-connection analysis between the satellite-derived ET time series and other related variables, such as NDVI, NDWI, NDDI and precipitation, was performed. Furthermore, the Stream Flow (SF) time series of the river and the dam that supply water to the irrigation scheme, namely, the Keiskamma River, Sandile Dam (monthly volumes of the dam level) were also analysed. In general, the results of this research indicate a significant upwards trend of ET, which is indicated by a z-score of approximately +3.898 over the period of 18 years. On the other hand, there is significant negative decrease on other variables, such as precipitation (z-score = -2.6134), and NDDI (z-score = -9.886), over the 18-year study period. Through the use of multi-linear regression analyses, it was observed that there is a statistically significant relationship between ET and NDVI (p-value < 7.89×10^{-11}), SF (p-value = 2.32×10^{-06}), Precipitation (p-value < $2x10^{16}$) and NDDI (p-value = 0.0208), respectively. This implies that, ET strongly depends on the precipitation, vegetation health and the response of the vegetation to drought conditions. Both the BFAST method and SQ-MK showed that the trend of the ET time series started to turn to positive (upwards) with significant direction just after the year 2010, while precipitation indicates signs of downward, significant trends. Moreover, the Wavelet analysis of the ET time series was performed to identify periodicities that are in the signal. The Wavelet Coherence Method was used to assess the tele-connection between ET and Precipitation. Not surprisingly, both the precipitation time series and the ET time series recorded a significant seasonality peak, with several short periodicities earlier in the signal. In terms of tele-connections, the Wavelet Coherence Spectrum indicated a strong seasonal tele-connection between precipitation and ET, with a cross-spectral phase indicating that precipitation is always leading to changes in ET.

ARIMA, ANN and Hybrid ARIMA-ANN models were trained using ET time series data for 15 years (2001 - 2015), and the remaining time series data for 3 years (2015 - 2018) was used as a test set for the model's performance. The performance of these models was assessed using four modelling accuracy measures: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and the Correlation Coefficient (R). In general, the three models are able to produce good predictions of ET. According to the comparison of the model performance statistics, it can be concluded that the Hybrid (ARIMA-ANN) model assures a reliable ET prediction for Keiskammahoek Irrigation Scheme. The model outperformed other models, with lower values: RMSE = 33.80, MAR = 27.02, MAPE = 17.31 and R = 0.94. This indicates that the combined model of ARIMA and ANN is a better option as the hybrid models are able to capture both linearity and non-linearity in the time series of ET which, in turn, produces better results. After the best model was selected from the three models used, the Hybrid (ARIMA-ANN) model was used to predict ET for 5 years (2018 to 2023).

5.2 CONCLUSION

This study uses used 18 years' time series data from (2001 to 2018) collect from Keiskammahok Irrigation Scheme in Eastern Cape South Africa using machine learning with its main objective as to predict Evapotranspiration. The Three prediction models, (ARIMA, ANNs and Hybrid) which has proved over the years to be efficient in predicting ET were employed. ARIMA models has been appraised by many researchers with as a capable model to predict the linear time series whilst ANNs has consecutively been praised with its ability of predicting a non-linear time serios. The third model Hybrid (ARIMA-ANN) was also employed in order to find best results which include both linearity and non-linearity.

After three models were employed, the hybrid models proved its supiority over the two ARIMA and ANN models with four model performance accuracies (RMSE, MAPE, MAE and Correlation Statistics) proving the Hybrid (ARIMA-ANNs) as the promising model that can predict ET for Keikammahoek Irrigation Scheme.

5.3 RECOMMENDATIONS

The consistent increase in ET and decrease of precipitation and other variables like streamflow and Sandile Dam levels is a concern as this indicate signs of drought in the study area. It is therefore recommended that drought in this community. It is strongly recommended that accurate time series prediction studies, similar to this one, are done as well as drought assessment studies.

REFERENCES

Abdel-Hamid, A., Dubovyk, O., Graw, V., & Greve, K. (2020). Assessing the impact of drought stress on grasslands using multi-temporal SAR data of Sentinel-1: a case study in Eastern Cape, South Africa. *European Journal of Remote Sensing*.

Abdullahi, J., & Elkiran, G. (2018). Prediction of the future impact of climate change on reference evapotranspiration in Cyprus using artificial neural network. *Science direct, Procedia Computer Science*, 276-283.

Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Comparison of ARIMA and Artificial Neural Networks Modelsfor Stock Price Prediction. *Journal of Applied Mathematics*.

Adler, J., & Ingela, P. (2010). Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry Part 1*, 733-742.

Adler, J., & Part, A. (2010). Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Crytometry Part A*, 733-742.

Affari, Department of Environmental, GIBB, & South African National Road Agency SOC ltd. (2016). i*mprovement of national route n2 section 14 & 15 from green river (km 60.0) to zwelitsha intersection (km 6.00) & the new breidbach interchange (km 9.8).* King Williamns Town: South African Road Agency.

Aksoy, S., Gorucu, O., & Sertel, E. (2019). Drought Monitoring using MODIS derived indices and Google Earth Engine Platform. *2019 8th International Confrence on Agro geoinformatics*.

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage. *FAO, Rome*.

Anapalli, S. S., Fisher, D. K., Reddy, K. N., Rajan, N., & Pinnamaneni, S. R. (2019). Modeling evapotranspiration for irrigation water management in a humid climate climate. *Elsevier, Agricultural Water management*.

Anderson, D., & McNeill, G. (1992). *Artificial Neuralnetworks Technology.* Utica, New York: A DACS State of the Art report.

Anderson, O. D. (1995). More effective time-series analysis and forecasting . *Journal of Computational and Applied Mathematics 64*, 117-147.

Annandale, J. G., Stirzaker, R. J., Singels, A., Laan, M. V., & Laker, M. C. (2011). Irrigation scheduling research: South African experiences and future prospects. *Water Reaserch Commision 40-Year Celebration Conference* (pp. 751-764). Water SA.

Apraku, A., Akpan, W., & Moyo, P. (2019). Indegenous Knowledge, Global Ignorance?, Insights from an Eastern Cape climate Change Study. *South African Review of Sociology*.

Archer, E., Du Toit, J., Engelbrecht, C., Hoffman, M. T., Landman, W., Malherbe, J., & Stem, M. (2022). The 2015-19 multi year drought in the Eastern Cape, South Africa: it's evolution and impacts on agriculture. *Journal of Arid Environment*.

Asadi, A., Vahdat, S. F., & Sarraf, A. (2013). The forecasting of Potential Evapotranspiration using time series analysis in humid and semi humid regions. *American Journal of engineering Research*, 296-302.

Aslanargun, A., Mammadov, M., Yazici, B., & Yolacan, S. (2007). Comparison of ARIMA, neural networks and hybridmodels in time series: tourist arrival forecasting. *Journal of Statistical Computation and Silumation*, 29-53.

Averbeke, W. V., M'Marete, C. K., Igodan, C. O., & Belete, A. (1998). *An investigation into food plot production at irrigation schemes in central Eastern CApe.* Eastern Cape: Water Research Commision Report.

Ayub, S., & Jafri, Y. Z. (2020). Comparative Study of an ANN-ARIMA Hybrid Model for Predicting Karachi Stock Price. *American Journal of Mathematics*, 1-9.

Bachour, *R. (2013). Evapotranspiration Modeling and Forecasting for Efficient Management of Irrigation Command Areas.* Loga, Utah: Utah State University.

Bari, S. H., Rahman, M. T., Hussain, M. M., & Ray, S. (2015). Forecasting Monthly Precipitation in Sylhet City Using ARIMA Model. *Civil and Nevironmental Research*.

Bates, J. M., & Clive, G. W. (1969). The combination of forecast. *Journal of the operational research society*, 451-468.

Bencherif, H., Toihir, A. M., Mbatha, N., Sivakumar, V., Du Prezz, D. J., Begue, N., & Coetzee, g. (2020). Ozone Variability and Trend Estimates from 20-Yearsof Ground-Based and Satellite Observations at IreneStation, South Africa. *atmosphere*.

Bhawna, & Gobind. (2015). Research Methodology and Approaches. *JOurnal of Research & Method in Education*, 48-51.

Birylo, M., Rzepeck, Z., Kuczynska-Siehien, J., & Nastula, J. (2018). Analysis of water budget prediction accuracy usingARIMA models. *Water Science & Technology: Water Supply*.

Botai, C. M., Botai, J. O., Adeola, A. M., Wit, J. P., Ngcongwane, K. P., & Zwane, N. N. (2020). Drought Risk Analysis in the Eastern Cape Province of South Africa: The Copula Lens. *Water*.

Bouznad, I.-E., Guastaldi, E., Zirulia, A., Brancale, m., Barbagli, A., & Bendusmia, D. (2020). Trend analysis and spatiotemporal prediction of precipitation, temperature, and evapotranspiration values using the ARIMA models: case of the Algerian Highlands. *Arabian Journal of geosciences*.

Brown, D. L. (1969). *The study of the animal and crop production systems and potential of the bantu ciskeian territories.* Natal.

Burns, P. (2002). Robustness of the Ljung-Box Test and its Rank Equivalent . *SSRN*.

Calzadilla, A., Zhu, T., Rehdanz, K., Tol, R. S., & Ringler, C. (2014). Climate change and agriculture: Impacts and adaptation options in South Africa. *Elsevier, Water resources and economics*, 24-48.

Caminha, H. D., da Silva, T. C., & da Rocha, A. R. (2017). Estimating Reference Evapotranspiration using Data Mining Prediction Models Models and Feature Selection. *ICEIS*, 272-279.

Campos, J. C., Sillero, N., & Brito, J. C. (2012). Normalized difference water indexes have dissimilar performances indetecting seasonal and permanent water in the Sahara–Sahel transition zone. *Journal of Hydrology*, 438-446.

Chahine, M. T. (1992). The hydrological cycle and its influence on climate. *Nature Publishing Group*.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? –Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 1247-1250.

Chapungu, L., & Nhamo, L. (2016). An Assessment of the Impact of Climate Change on Plant Species RichnessThroughanAnalysis of the Normalised Difference Water Index(NDWI)in Mutirikwi Sub-catchment, Zimbabwe. *South African Journal of Geomatics*.

Chatfield, C. (2000). *Time-Series Forecasting.* Chapman and Hall/CRC.

Chattopadhyay, S., & Chattopadhyay, G. (2010). Univariate modelling of summer-monsoon rainfall time series: Comparison between ARIMA and ANN. *Comtes Rendus Geoscience*, 100-107.

Chen, P., Yuan, H., & Shu, X. (2008). Forecasting Crime Using the ARIMA Model. *Fifth International Conference on Fuzzy Systems and Knowledge Discovery.* IEEE.

Chersich, M. F., Wright, C. Y., Venter, F., Rees, H. S., & Erasmus, B. (2018). Impact of Climate change on Health and Wellbeing in South Africa. *international journal of environmental research and Public Health*.

Cihlar, J., laurent, L. S., & Dyer, J. A. (1991). Relation between the normalized difference vegetation index and ecological variables. *Remote sensing of Environment*, 279-298.

Clemen, R. T. (1989). Combinging forecasts: A review and annotated bibliography. *International journal of forecasting*, 559-583.

Connor, J. D., Schwabe, K., King, D., & knapp, K. (2012). Irrigated agriculture and climate change: The influence of water supply variability and salinity on adaption. *Elsevier, Ecological Economics*, 149-157.

Contreras, J., Espinola, R., Nogales, F. J., & Conejo, A. J. (2003). ARIMA Models to PredictNext-Day Electricity Prices. *IEEE Transactions on Power Systems*.

Dang, X., Peng, H., Wang, X., & Zhang, H. (2008). Theil-Sen Estimators in a Multiple Linear Regression Model. *Olemiss Edu*.

De, M., Golden, A., Le, G. B., & Rossi, F. (2016). Mean absolute percentage error. *Neuricomputing*, 38-48.

Diaz-Robles, L. A., Ortega, J. C., Fu, J. S., Reed, G. D., Chow, J. C., Watson, J. G., & Mondada-Herrera, J. A. (2008). A hybrid ARIMA and artificial neural networks model to forecastparticulate matter in urban areas: The case of Temuco, Chile. *Athmospheric Environment*, 8331-8340.

Diks, C. G., & Vrugt, J. A. (2010). omparison of point forecast accuracy of model averagingmethods in hydrologic applications. *Stoch Environ Res Risk Assess*, 809-820.

Dingwell, J. B. (2000). Nonlinear time series analysis of normal and pathological human walking. *American Intsitute of Physics*.

Dinpashoh, Y., Jhajharia, D., Fakheri-Fard, A., Singh, V. P., & Kahya, E. (2011). Trends in reference crop evapotranspiration over Iran. *Journal of Hydrology*, 422-433.

Du, T. L., Bui, D. D., Nguyen, M. D., & Lee, H. (2018). Satellite-Based, Multi-Indices for Evaluation ofAgricultural Droughts in a Highly DynamicTropical Catchment, Central Vietnam. *Water*.

Dutta, B., Smith, W. N., Grant, B. B., Pattey, E., Desjardins, R. L., & Li, C. (2016). Model development in DNDC for the prediction of evapotranspiration and water use in temperate field cropping systems. *Enviromental modelling Software*, 9-25.

Elwasify, A. I. (2015). A Combined Model between Artificial Neural Networks and ARIMA Models. *International Journal of Research in Commerce Economics and Management*, 134-140.

Fanadzo, M., Chiduza, C., Mnkeni, P., Stoep, I. d., & Stevens, J. (2010). management practices as a cause for low water productivity at Zanyokwe Irrigation Scheme. *Water SA*.

Fattah, J., Ezzine, L., Aman, Z., Moussami, H. E., & Lachhab, A. (2018). Jamal. *Forecasting of demand usingARIMA model*.

Feng, G., Cobb, S., Abdo, Z., Fisher, D. K., Ouyang, Y., Adeli, A., & Jenkins, J. N. (2016). Trend Analysis and Forecast of Precipitation, Reference Evapotranspiration, andRainfall Deficit in the Blackland Prairie of Eastern Mississippi. *Journal of applied meteology and climatology*.

Freeman, E. A., & Moisen, G. (2008). PresenceAbsence: AnRPackage for PresenceAbsence Analysis. *Journal of Statistical Software*.

Gandrud, C. (2018). *Reproducible Research with R and RStudio Second Edition.* Chapman and Hall/CRC.

Gao, B.-C. (1996). NDWI A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space. *Elsevier Science*, 257-266.

Gautam, R., & Sinha, A. (2016). Time series analysis of reference crop evapotranspiration for Bokaro District, Jharkhand, India. *Journal of Water and Land Development*, 51-56.

Gharde, K. D., Kothari, M., & Mahale, D. M. (2016). Developed Seasonal ARIMA Model to Forecast Streamflow for Savitri Basin in Konkan Region of Maharshtra on Daily Basis. *J.Indian Soc Coastal Agric*, 110-119.

Ghorbani, M. A., Kazempour, R., Chau, K.-W., Shamshirband, S., & Ghazvinei, P. T. (2018). tcfm20Forecasting pan evaporation with an integratedartificial neural network quantum-behavedparticle swarm optimization model: a case study inTalesh, Northern Iran. *Engineering Applications of Computational FluidMechanics*.

Gorantiwar, S. D., Meshram, D. T., & Mittal, H. K. (2011). Seasonal ARIMA model for generation and forecasting evapotranspirtion of Solapur district of Maharashtra. *Jounrla of Agrometeorology*, 119-122.

Gorelick, N., Hancher, M., Dixon, M., Llyuschenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 18-27.

Gouveia, C. M., Bastos, A., Trigo, R. M., & DaCamara, C. C. (2012). Drought impacts on vegetation in the pre- and post-fire events overIberian Peninsula. *Natural Hazards and Earth Systems Science*, 3123-3237.

Granger, C. W. (1984). Improved methods of Combining Forecast. *Journal of Forecasting*, 197-204.

Graw, V., Ghazaryan, G., Dall, K., Gomez, A. D., Abdel-hamid, a., Jordaan, A., . . . Dubovyk, O. (2017). Drought Dynamics and Vegetation Productivity in Different Land Management Systems of Eastern Cape, South Africa—A Remote Sensing Perspective. *Sustainability*.

Grinsted, A., Moore, J. C., & Jevrejeva, S. (2004). Application of the cross wavelet transform and waveletcoherence to geophysical time series. *Nonlinear Processes in Geophysics, European Geosciences Union*, 561-566.

Gulacsi, A., & Kovacs, F. (2018). Drought monitoring of forest vegetation using MODIS-based normalized difference drought index in HungaryAndrás GULÁCSI1andFerenc KOVÁCS1AbstractIn this paper, several spectral indices based on spectral reflectance data from the Terra satellite's Moderate . *Hungarian Geographical Bulletin*.

Haindongo, P. N. (2009). *An Investigation of the Factors Influencing Vegetation Stress in a part of the Keiskamma Catchment, Eastern Cape: a Remote Sensing and GIS Approach.* Port ELizabeth: Nelson Mandela University.

Hamdi, M. R., Bdour, A. N., & Tarawneh, Z. S. (2008). Developing Reference Crop Evapotranspiration Time Series Simulation Model Using Class a Pan: A Case

Study for the Jordan Valley/ Jordan. *Jourdan Journal of Earth and ENvironmental Science*, 33-44.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutoria. *Statistical Science*, 382-417.

Hsiao, T. C., Steduto, P., & Fereres, E. (2007). A systematic and quantitative approach to improve water use efficiency in agriculture. *Irrigation Sci*, 209-231.

Hussain, s. (2015). Educational Data Mining Using R Programming and R Studio. *Journal of applied and fundamental science*.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting principles and practice.* Otexts.

Jain, K. A., Mao, J., & Mohiuddin, M. K. (1996). Artificial neural networks: A tutorial. *Computer*, 31-44.

Jasechko, S., Sharp, z. D., Gibson, J. G., Birks, S. J., Yi, Y., & fawcett, p. J. (2013). Terrestrial water fluxes dominated by transpiration. *Nature*, 347-350.

Jha, G. K. (2007). Artificial Neural Networks and its Applications. *Artificial Neural networks and its applications*.

Jovanovic, N., Garcia, C., Bugan, R. D., Teich, I., & Rodriguez, C. M. (2014). Validation of remotely-sensed evapotranspiration and NDWI using ground measurements at Riverlands, South Africa. *Water SA*.

Jovanovic, N., Mu, Q., Bugan, R. D., & Zhao, M. (20215). Dynamics of MODIS evapotranspiration in South Africa Nebo. *Water SA*.

Jury, M. R., & levey, K. (1993). The Eastern Cape drought. *Water SA*.

kahsay, G. a., & Hansen, L. G. (2016). The effect of climate change and adaption policy on agricultural production in eastern Africa. *Elsevier, Ecological Economics*, 54-64.

Kambale, J. B., Singh, D. K., & Sarangi, A. (2017). Impact of Climate Change on Ground Water Recharge in a Semi-Arid Region of Northern India. *Impact of climate change on groundwater recharge in semi-arid region of Northen India*.

Karbasi, M. (2018). Forecasting of Multi-Step Ahead ReferenceEvapotranspiration Using Wavelet- Gaussian ProcessRegression Model. *Water resour Manage*, 1035-1052.

Khair, U., Fahmi, H., Hakim, S., Al, & Rahim, R. (2017). Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error.

International Conference on Information and Communication Technology. IOP Publishing.

Khandelwal, I., Adhikari, R., & Verma, G. (2015). Time Series Forecasting using Hybrid ARIMA and ANN Models based on DWT Decomposition. *Sciencedirect, Procedia Computer Science*, 173-179.

Khanna, R., Piyush, & Bhalla, P. (2014). Study of Artificial Neural Network. *International Journal of Research in Information Technology*, 271-276.

Khoshhal, J., & Mokarram, M. (2012). Model for Prediction of Evapotranspiration Using MLP Neural Network. *International Journal of Environmental Sciences, 3*.

Kihoro, M. J., Otieno, O. R., & Wafula, C. (2004). Seasonal Time Series Forecasting: a Comparative Study of ARIMA and ANN Models. *African Journal of Science and Technology*, 41-49.

Kim, Y.-O., Jeong, D., & Ko, I. H. (2006). Combining Rainfall-Runoff Model Outputs for ImprovingEnsemble Streamflow Prediction. *Journal of Hydrologic Engineering* .

Kishore, V., & Pushpalatha, M. (2017). Forecasting Evapotranspiration for Irrigation Scheduling using Neural Networks and ARIMA. *International Journal of Applied Engineering Research*, 10841-10847.

Koutsoyiannis, D. (2020). Revisiting global hydrological cycle: Is it intensifying? *Hydrology and Earth System Sciences*.

Kumar, M., Raghuwanshi, N. S., SIngh, R., Wallender, W. W., & Pruitt, W. O. (2002). Estimating Evapotranspiration using Artificial NeuralNetwork. *Journal of Irrigation and Drainage Engineering*.

Kumar, S., Tanwar, S. P., & Singh, A. (2017). *Drought Mitigation and Management.* Jodhpur: Scientific Publishers.

Lee, S.-J., Cho, J., Hong, S., Ha, K.-j., Lee, H., & Lee, y.-W. (2016). On the relationships between satellite-baseddrought index and gross primary production in theNorth Korean croplands, 2000–2012. *Remote sensing*, 790-799.

Lee, S.-J., Cho, J., Hong, S., Ha, K.-J., Lee, H., & Lee, Y.-W. (2016). On the relationship between satellite-based drought index gross primary production in the North koren croplands, 2000-2012. *Remote sensing letters*, 790-799.

Loua, R. T., Bencherif, H., Nkanyiso, M., Begue, N., Haunchecome, A., Bamba, Z., & Sivakumar, V. (2019). Study on Temporal Variations of Surface Temperatureand Rainfall at Conakry Airport, Guinea: 1960–2016. *Climate*.

Ma, M., & Wang, Z. (2019). Prediction of the Energy Consumption Variation Trend in South Africa based on ARIMA, NGM and NGM-ARIMA Models. *Energies*.

Machida, F., Andrzejak, A., Matias, R., & Vicente, E. (2013). On the Effectiveness of Mann-Kendall Test forDetection of Software Aging. *IEEE International Symposium on Software Reliability Engineering Workshops*.

Mahlalela, P. T., Blamey , R. C., Hart, N. C., & Reason, C. J. (2020). Drought in the Eastern Cape region of South Africa and trends in rainfall characteristics. *Climate Dynamics*.

Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of waterresources variables: a review of modelling issues and applications. *Environmental Modelling & Software*, 101-124.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., . . . Winkler, R. (1982). The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition. *Journal of Forecasting*, 111-153.

Mallikarjuna, M., & Rao, R. P. (2019). Evaluation of forecasting methods fromselected stock market returns. *Financial Innovation*.

Mandleni, B., & Anim, F. (2011). Climate Change Awareness and decision on Adaption Measures by Livestock famers. *85rd Annual confrence of the Agriculture Economics Society.* Florida.

Maselli, F., Rodolfi, A., Botta, L., & Conese, C. (1996). Evaluation of forest fire risk by the analysis of environmental data and TM images. *International Journal of Remote Sensing*, 1417-1423.

Mbatha, N., & Bencherif, H. (2020). Time Series Analysis and Forecasting Using a NovelHybrid LSTM Data-Driven Model Based on EmpiricalWavelet Transform Applied to Total Column of Ozoneat Buenos Aires, Argentina (1966–2017). *atmosphere*.

Mbatha, N., & Xulu, S. (2018). Time Series Analysis of MODIS-Derived NDVI forthe Hluhluwe-Imfolozi Park, South Africa: Impact of Recent Intense Drought. *Climate*.

Mingjun, D., Yili, Z., Linshan, L., Wei, Z., Zhaofeng, W., & Wanqi, B. (2007). The relationship between NDVI and precipitation on the Tibetan plateau. *Journal of geographical Science*.

Mishra, A. K., & Singh, V. P. (2011). Drought modeling – A review. *Elsevier*, 157-175.

Mondal, P., Shit, L., & Goswami, S. (2014). Study of Effectiveness of Time Series Modeling (ARIMA) in Forecasting Stock Prices. *International Journal of Computer Science, Enfgineering and Applications*.

Montheith, J. L. (1965). Evaporation and environment. *Symposia of the Society for experimental Biology*.

Montheith, J. L. (1981). Evaporation and surface temperiture. *Quartely Journal of the Royal Meteorological Society*, 1-27.

Morimoto, T., Ouchi, Y., Shimizu, M., & Baloch, M. S. (2007). Dynamic optimization of watering Satsuma mandarinusing neural networks and genetic algorithms. *Agricultural Water Management*, 1-10.

Muijs, D. (2004). *Doing quantitative research in education: With SPSS.* New Delhi, London: Sage.

Mukaka, M. M. (2012). Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal*, 69-71.

NG, E. K., & Chan, J. C. (2012). Geophysical Applications of Partial Wavelet Coherence and MultipleWavelet Coherence. *Journal of atmospheric and Oceanic Technology*, 1845-1853.

Ngaka, M. J. (2012). Drought preparedness, impact and response: A case of eastern Cape and Free State provinces of South Africa. *Journal of disaster Risk Studies*.

Nkondo, M. N., zyl, F. V., Keuris, N., & Shrener, b. (2012). Proposed national water resources strategy Summary. *Managing water for equitable and sustainability future*.

Ogunrinde, A. T., Oguntunde, P. G., Fasinmirin, J. T., & Akinwumiju, A. S. (2020). Application of artificial neural network for forecastingstandardized precipitation and evapotranspiration index:A case study of Nigeria. *Engineering Report*.

Olivier, F. C., & Singels, A. (2015). Increasing water use efficiency of irrigated sugarcane production in South Africa through better agronomic practices. *Field Crops research*, 87-98.

Orimoloye, I. R., Ololade, O. O., Mazinyo, S. P., Kalumba, A. M., Ekundayo, O. Y., Busayo, E. T., . . . Nel, W. (2019). Spatial assessment of drought severity in Cape Town area, South AfricaI.R. Orimoloyea,b,*, O.O. Ololadea, S.P. Mazinyob, A.M. Kalumbab, O.Y. Ekundayob,E.T. Busayob, A.A. Akinsanolac,d, W. NelbaCentre for Environmental Management, University of the Free St. *Heliyon*.

Palacios-Oruetes, A., Khanna, S., Litago, J., whiting, M. L., & Ustin, S. L. (2006). Assessment of NDVI and NDWI spectral indices using MODIS time series analysis and development of a new spectral index based on MODIS shortwave infrared bands. *international Conference of Remote Sensing and Geoinformation Processing*, (pp. 207-209). Trier.

Pandorfi, H., Bezerra, A. C., Atarassi, R. T., Vieira, F. M., Barbosa Filho, J. A., & Guiselini, C. (2016). Artificial neural networks employment in the predictionof evapotranspiration of greenhouse-grown sweet pepper. *Revista Brasileira deEngenharia Agrícola e Ambiental*, 507-512.

Parasuraman, K., Elshorbagy, A., & Carey, S. K. (2007). Modelling the dynamics of the evapotranspirationprocess using genetic programming. *Hydrological Science Journal*.

Patel, J. N., & Balve, P. N. (2016). Evapotranspiration Estimation with Fuzzy Logic. *IRES International Conference.* Zurich, Swizaerland: Int J Adv Mech Civ Eng.

Penman, H. L. (1956). Evaporation: An Introductory Survey. *Netherlands Journal of Agriculture Science*, 9-29.

Perret, S. R. (2002). 1 Water policies and smallholding irrigation schemesin South Africa: a history and new institutional challenges. *Journal Water Policy*, 283-300.

Pravilovic, S., & Appice, A. (2013). The intelligent forecasting model of time series. *Automation, Control and Intelligent Systems*, 90-98.

Psilovikos, A., & Elhang, M. (2013). Forecasting of Remotely Sensed Daily EvapotranspirationData Over Nile Delta Region, Egypt. *Water Resour Manage*, 4115-4130.

Qi, J., Du, J., Siniscalchi, S. M., Ma, X., & Lee, C.-H. (2020). On Mean Absolute Error for Deep Neural NetworkBased Vector-to-Vector Regression. *IEEE Signal Processing Letters*.

Ramanathan, V., Crutzen, P. J., Kiehl, J. T., & Rosenfeld, D. (2001). Aerosols, Climate, and the Hydrological Cycle. *Science compass*.

Ramoelo, A., Majozi, N., Mathieu, R., Jovanovic, N., Nickless, A., & Dzikiti, S. (2014). Validation of Global Evapotranspiration Product (MOD16) using Flux Tower Data in the African Savanna, South Africa. *Remote sensing*, 7406-7423.

Reza, D. A., & Debnath, T. (n.d.). An Approach to Make Comparison of ARIMA and NNAR Models For Forecasting Price of Commodities. 2020.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., . . . Da Silver. (2011). MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*.

Saayman, A., & Saayman, M. (2008). 96Determinants of inbound tourism to South Africa. *Tourism*, 81-96.

Santos, C., Lorite, I. J., Tasumi, M., Allen, R. G., & Fereres, E. (2007). Integrating satellite-based evapotranspiration with simulation models for irrigation management at the scheme level. *Agricultural science*, 277-288.

Schlesinger, W. H., & Jasechko, S. (2014). Transpiration in global water cycle. *Agricultural and forest meteorology*, 115-117.

Schultz, A., Wieland, R., & Lutze, G. (2000). Neural networks in agroecological modelling— stylish application or helpful tool? *Computers and Electronics in Agriculture*, 73-97.

Shang, H. L., & Booth, H. (2020). Synergy in fertility forecasting:improving forecast accuracy through modelaveraging. *Genus 76*, 1-23.

Shinde, P. P., Oza, K. S., & Kamat, R. K. (2017). Big Data Predictive Analysis:Using R Analytical Tool. *International confrerence on I-SMAC* (pp. 839-842). SOcial, Mobile, Analytics and Cloud.

Shuttleworth, W. J. (1979). *Evaporation.* Institute of hydrology.

Simons, D., & Laryea, S. (2005). Testing the efficiency of Selected African Stock Markets. *SSRN 874808*.

Simulator. (2005). *Overview of Different types of Forecasting Techniques.* Real options valuations.

Singels, S., & Smith, M. T. (2006). Provision of Irrigation Scheduling Advice to Small-Scale Sugarcane Farmers Using a Web-based Crop Model and Cellular Technology: a South African Case Study. *Irrigation and Drainage*, 363-372.

Sneyers, R. (1991). *On the statistical analysis of series of observations.*

Soh, Y. W., Koo, C. H., Huang, Y. F., & Fung, K. F. (2018). Application of artificial intelligence models for the prediction ofstandardized precipitation evapotranspiration index (SPEI) at Langat RiverBasin, Malaysia. *Computers and Electronics in Agriculture*, 164-173.

Sultana, S. R., Ali, A., Ahmad, A., Mubeen, M., UI-Haq, M. Z., Ahmad, S., . . . Jaafar, H. Z. (2014). Normalized Difference Vegetation Index as a Tool for WheatYield Estimation: A Case Study from Faisalabad, Pakistan. *The Scientific World Journal*.

Tavazohi, E., & Nadoushan, M. A. (2018). Assessment of Drought in the Zayandehroud Basin During 2000–2015 using nddi and spi indices. *Fresenius Environmental Bulletin*, 2332-2340.

Tawegoun, R., Belbrahem, R., & Chasseriaux, G. (2004). Modeling Evapotransplra tion Prediction on Nursery Area Using Recurrent Neural Network. *Fifth International Workshop on Artificial Intelligence in Agriculture* (pp. 89-91). Cairo: IFAC Proceedings .

Todaro, M. P., & Snith, S. C. (2012). Economic Development. *New York Library of congress cataloging*.

Torrence, C., & Compo, G. P. (1997). A Practical Guide to Wavelet Analysis. *Bulletin of American meteorology Society*.

Traore, S., Wang, Y.-M., & Kerh, T. (2008). Modeling Reference Evapotranspiration by Generalized Regression Neural Network in Semiarid Zone of Africa. *WSEAS Transactions on Information Science & Applications*.

TRAORE, S., WANG, Y.-M., & KERH, T. (2008). Modeling Reference Evapotranspiration by Generalized Regression Neural Network in Semiarid Zone of Africa . *WSEAS*, 991-1000.

Tsay, R. S. (2005). *Analyisis of Financial Time Series.* Hoboken, New Jersey: John Wiley & Sons. Inc.

Valipour, M. (2012). Ability of Box-Jenkins Models to Estimate of Reference Potential Evapotranspiration ,A Case Study: Mehrabad Synoptic Station, Tehran, Iran. *Journal of Agriculture and Veterinary Science*, 01-11.

Valipour, M., Banihabib, M. E., & Reza Behhahani, S. M. (2013). Comparison of the ARMA, ARIMA, and the autoregressive artificial neuralnetwork models in forecasting the monthly inflow of Dez dam reservoir. *Journal of Hydrology*, 433-441.

Verzani, J. (2002). *Simple R-Using R for Introductory Statistics.* New York: http://www. math. csi. cuny. edu/Statistics/R/simpleR/index.

Vicente-Serrano, S. M., Van der Schrier, G., Begueria, S., Azorin-Molina, C., & Lopez-Moreno, J.-I. (2015). Contribution of precipitation and reference evapotranspiration to drought 1indices under different climates. *Journal of Hydrology*, 42-54.

Wadi, S. A., Ismail, M. T., Alkhahazaleh, M. H., & Abdul Karim, S. A. (2011). Selecting Wavelet Transforms Model in Forecasting Financial Time Series Data Based on ARIMA Model. *Applied Mathematical Science*, 315-326.

Weisang, G., & Awazu, Y. (2008). Vagaries of the Euro: An Introduction to ARIMA Modeling. *Case studies in Business, Industry and government Statistics*, 45-55.

Wheeler, T. D., & Stroock, A. D. (2008). The transpiration of water at negative pressure in a synthethic tree. *Nature*, 208-212.

Widowati, Putro, S. P., Koshio, S., & Oktaferdian. (2016). Implementation of ARIMA Model to Asses Seasonal Variability Macrobenthic Assembages. *2nd International Symposium on Aquatic Processing and Health*, 277-284.

Wilfried, b. (2005). *Hydrology: an introduction.* Cambridge University press.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) overthe root mean square error (RMSE) in assessingaverage model performance. *Climate Research*, 79-82.

Xulu, S., Peerbhay, K., Gebreslaie, M., & Ismail, R. (2018). Drought Influence on Forest Plantations in Zululand,South Africa, Using MODIS Time Series andClimate Data. *Forest*.

Yang, F., White, A. M., Michaelis, A. A., Ichii, K., Hashimoto, H., Votava, P., . . . Nemani, R. R. (2006). Prediction of Continental-Scale Evapotranspirationby Combining MODIS and AmeriFlux DataThrough Support Vector Machine. *IEEE Transactions on Geoscience and Remote Sensing*.

Yue, S., & Wang, C. (2004). The Mann-Kendall Test Modified by EffectiveSample Size to Detect Trend in Serially CorrelatedHydrological Series. *Water Resources Management*, 201-218.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMAand neural network model. *Neurocomputing*, 159-175.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks:The state of the art. *International Journal of Forecasting*, 35-62.

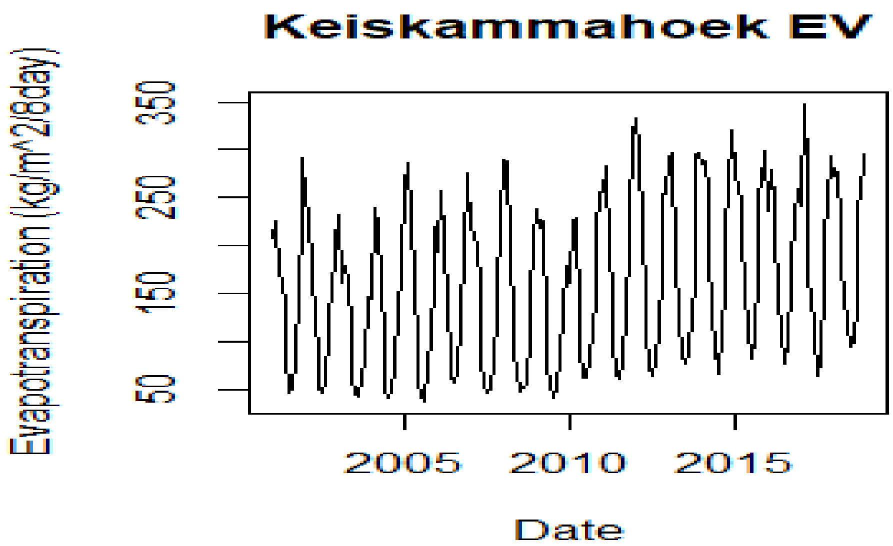
Zhang, L. L., & Wylie, B. (2009). Analysis of Dynamic Thresholds for the Normalized Difference Water Index. *Photogrammetric Engineering & remote Sensing*, 1307-1317.

Zhang, L., Zhang, G. X., & Li, R. R. (2016). Water Quality Analysis and Prediction Using Hybrid Time Series and Neural Network Models. *Journal of Agricultural Science and Technology*, 975-983.

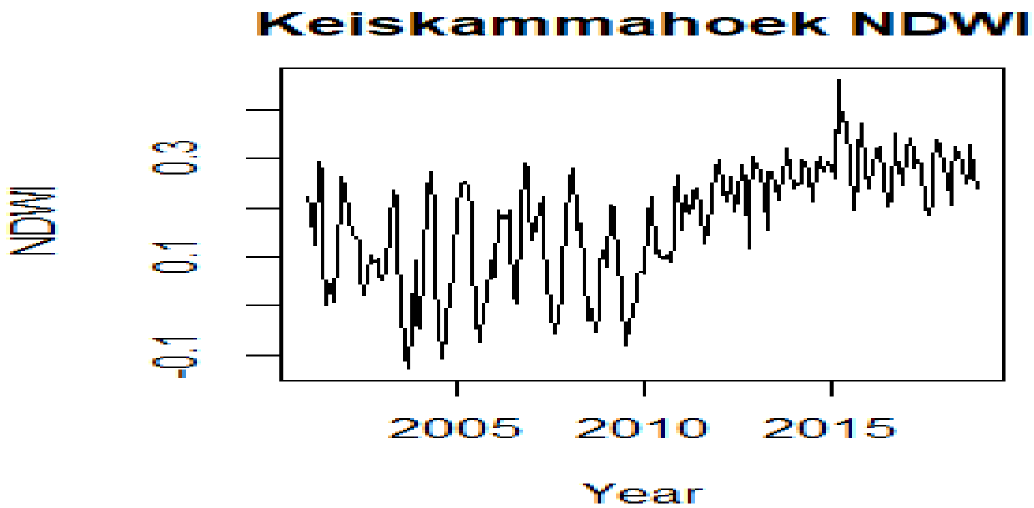
Ziervogel, G., New, M., Archer, G. V., Midgley, G., Taylor, A., Mamann, R., . . . Warburton, M. (2014). Climate change impacts and adaption in South Africa. *Climate Change*.

APPENDIX

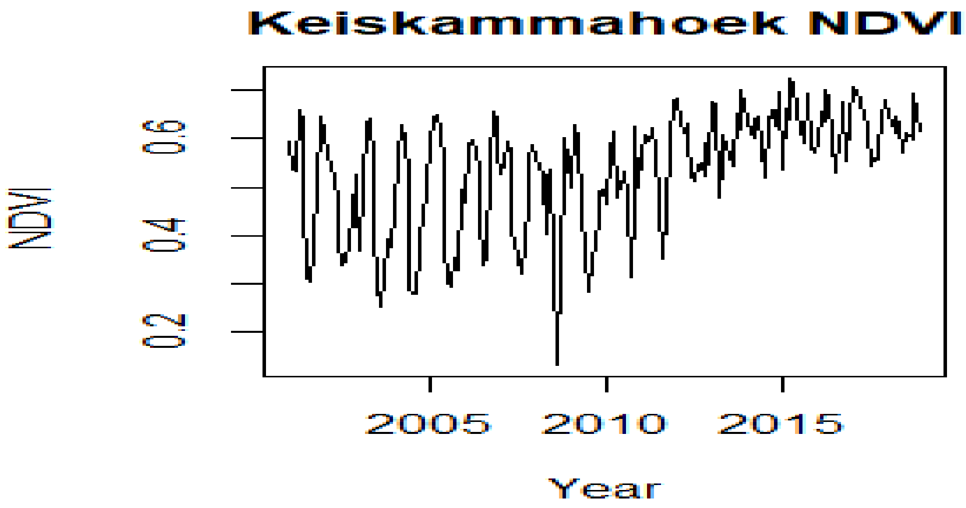
Appendix A :18 years' time series data for various variables at Keiskammahoek irrigation scheme.



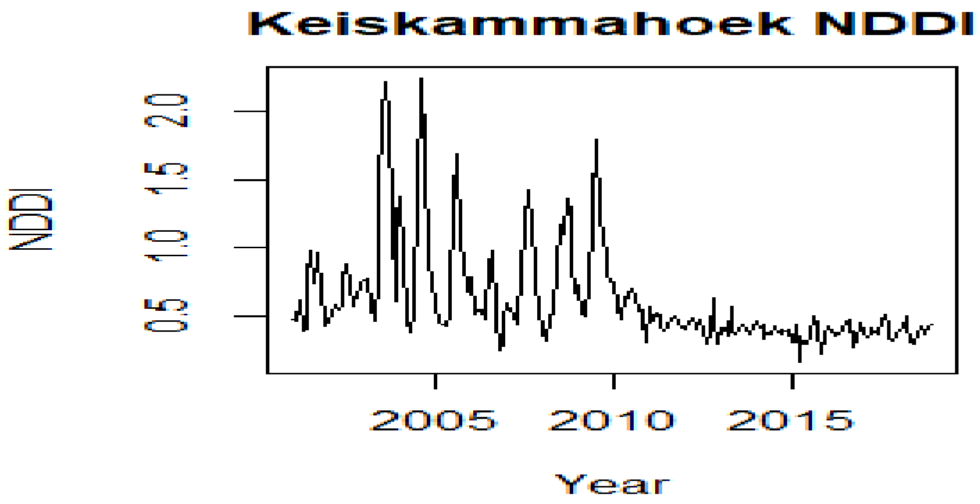
Appendix A.1 : ET time series data for 18 years (2001-2018).



Appendix A.2: NDWI time series data for 18 years (2001-2018).



Appendix A.3: NDVI time series data for 18 years (2001-2018).



Appendix A.4: Calculated NDDI drought index using NDVI and NWI for 18 years (2001-2018).